PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 :
A61K 39/29, 39/42, C07H 17/00
C07K 7/10, 15/04, C12N 7/02
C12Q 1/68, 1/70, G01N 33/531
G01N 33/536, 33/543

A1

(11) International Publication Number: WO 89/12462

(43) International Publication Date: 28 December 1989 (28.12.89)

(21) International Application Number: PCT/US89/02648

(22) International Filing Date: 16 June 1989 (16.06.89)

(30) Priority data:

208,997	17 June 1988 (17.06.88)	US
336,672	11 April 1989 (11.04.89)	US

(71) Applicants: GENELABS, INCORPORATED [US/US]; 505 Penobscot Drive, Redwood City, CA 94063 (US). THE GOVERNMENT OF THE UNITED STATES OF AMERICA as represented by THE SECRETARY OF THE DEPARTMENT OF HEALTH AND HUMAN SERVICES AND HIS SUCCESSORS [US/US]; Washington, DC 20231 (US).

(72) Inventors: REYES, Gregory, R. ; 2112 St. Frances Drive, Palo Alto, CA 94303 (US). BRADLEY, Daniel, W. ; 2938 Kelly Court, Lawrenceville, GA 30244 (US).

(74) Agent: NEELEY, Richard, L.; Leydig, Voit & Mayer, 350 Cambridge Avenue, Suite 200, Palo Alto, CA 94306 (US).

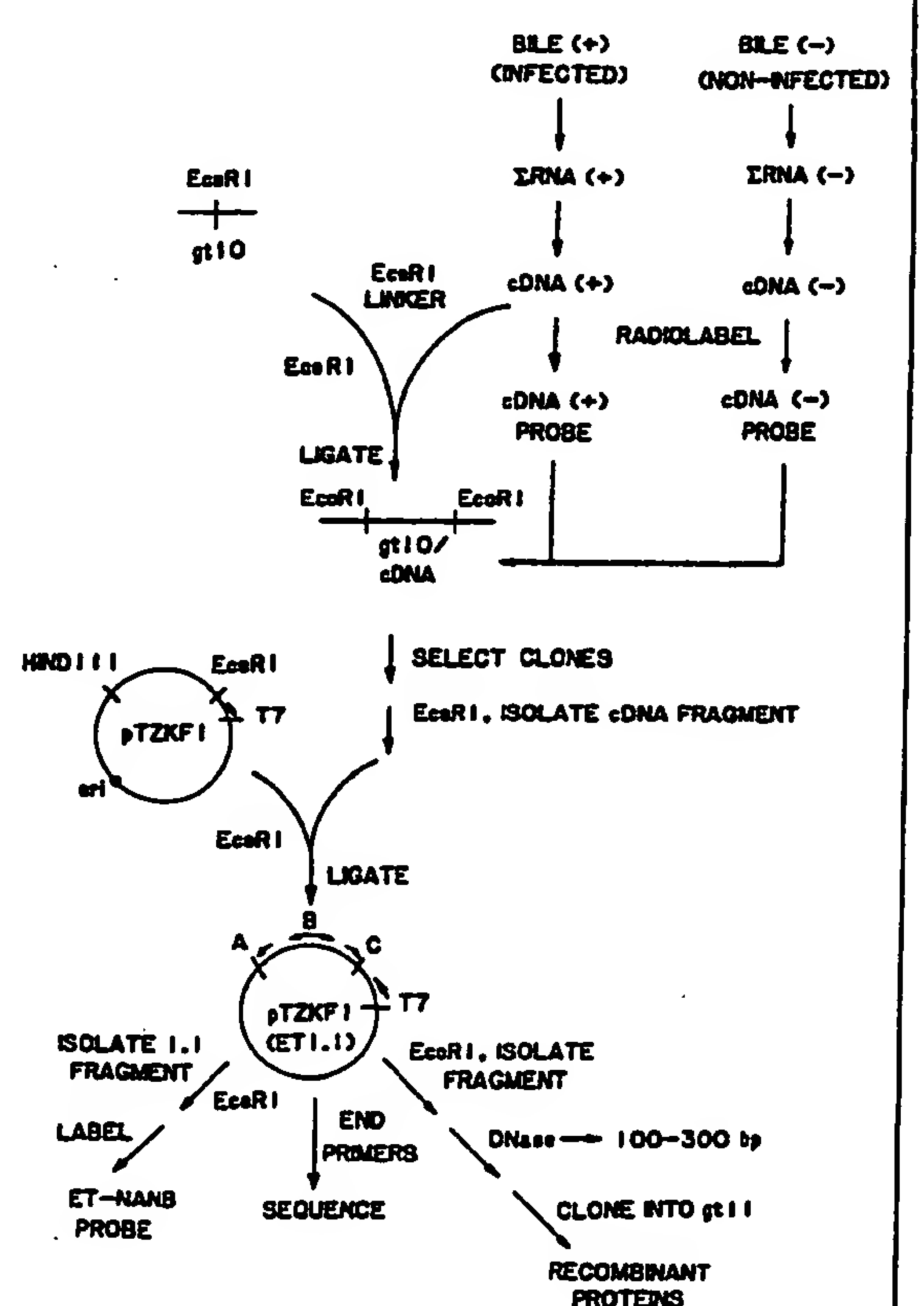
(81) Designated States: AT (European patent), AU, BE (European patent), CH (European patent), DE (European patent), DK, FR (European patent), GB (European patent), IT (European patent), JP, KR, LU (European patent), NL (European patent), SE (European patent).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: ENTERICALLY TRANSMITTED NON-A/NON-B HEPATITIS VIRAL AGENT

(57) Abstract

Viral proteins derived from an enterically transmitted non-A/non-B viral hepatitis agent are disclosed. In one embodiment, the protein is immunologically reactive with antibodies present in individuals infected with the viral hepatitis agent. This protein is useful in a diagnostic method for detecting infection by the enterically transmitted agent. Also disclosed are DNA probes derived from a cloned sequence of the viral agent. These probes are useful for identifying and sequencing the entire viral agent and for assaying the presence of the viral agent in an infected sample, using probe-specific amplification of virus-derived DNA fragments.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Fasso
BG	Bulgaria
BJ	Benin
BR	Brazil
CF	Central African Republic
CG	Congo
CH	Switzerland
CM	Cameroon
DE	Germany, Federal Republic of
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
HU	Hungary
IT	Italy
JP	Japan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
MC	Monaco
MG	Madagascar
ML	Mali
MR	Mauritania
MW	Malawi
NL	Netherlands
NO	Norway
RO	Romania
SD	Sudan
SE	Sweden
SN	Senegal
SU	Soviet Union
TD	Chad
TG	Togo
US	United States of America

ENTERICALLY TRANSMITTED NON-A/NON-B HEPATITIS VIRAL AGENT

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of U.S. Application Serial No. 336,672, filed April 11, 1989, which is a continuation-in-part of U.S. Application Serial No. 208,997, filed June 17, 1988, both of which are herein incorporated by reference.

INTRODUCTION

Field of Invention

This invention relates to recombinant proteins, genes, and gene probes and more specifically to such proteins and probes derived from an enterically transmitted nonA/nonB hepatitis viral agent, and to diagnostic methods and vaccine applications which employ the proteins and probes.

Background

Enterically transmitted non-A/non-B (ET-NANB) hepatitis viral agent is the reported cause of hepatitis in several epidemics and sporadic cases in Asia, Africa, and the Indian subcontinent. Infection is usually by water contaminated with feces, although the virus may also spread by close physical contact. The virus does not seem to cause chronic infection. The viral etiology in ET-NANB has been demonstrated by infection of volunteers with pooled fecal isolates; Immune electron microscopy (IEM) studies have shown virus particles with 27-34 nm diameters in stools from infected individuals. The virus particles reacted with

antibodies in serum from infected individuals from geographically distinct regions, suggesting that a single viral agent or class is responsible for the majority of ET-NANB hepatitis seen worldwide. No antibody reaction was seen in serum from individuals infected with blood-transmitted NANB virus, indicating a different specificity between the two NANB types.

In addition to serological differences, the two types of NANB infection show distinct clinical differences. ET-NANB is characteristically an acute infection, often associated with fever and arthralgia, and with portal inflammation and associated bile stasis in liver biopsy specimens (Arankalle). Symptoms are usually resolved within six weeks. Blood-transmitted NANB, by contrast, produces a chronic infection in about 50% of the cases. Fever and arthralgia are rarely seen, and inflammation has a predominantly parenchymal distribution (Khuroo, 1980). The two viral agents can also be distinguished on the basis of primate host susceptibility. ET-NANB, but not the blood-transmitted agent, can be transmitted to cynomolgus monkeys. The blood-transmitted agent is more readily transmitted to chimpanzees than is ET-NANB (Bradley, 1987).

There have been major efforts worldwide to identify and clone viral genomic sequences associated with ET-NANB hepatitis. One goal of this effort, requiring virus-specific genomic sequences, is to identify and characterize the nature of the virus and its protein products. Another goal is to produce recombinant viral proteins which can be used in antibody-based diagnostic procedures and for a vaccine. Despite these efforts, viral sequences associated with ET-NANB hepatitis have not been successfully identified or cloned heretofore, nor have any virus-specific proteins been identified or produced.

Relevant Literature

Arankalle, V.A., et al., The Lancet, 550 (March 12, 1988).

Bradley, D.W., et al., J Gen. Virol., 69:1 (1988).

Bradley, D.W. et al., Proc. Nat. Acad. Sci., USA, 84:6277 (1987).

Gravelle, C.R. et al., J. Infect. Diseases, 131:167 (1975).

Kane, M.A., et al., JAMA, 252:3140 (1984).

Khuroo, M.S., Am. J. Med., 48:818 (1980).

Khuroo, M.S., et al., Am. J. Med., 68:818 (1983).

Maniatis, T., et al. Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory (1982).

Seto, B., et al., Lancet, 11:941 (1984).

Sreenivasan, M.A.m et al., J. Gen. Virol., 65:1005 (1984).

Tabor, E., et al., J. Infect. Dis., 140:789 (1979).

SUMMARY OF THE INVENTION

Novel compositions, as well as methods of preparation and use of the compositions are provided, where the compositions comprise viral proteins and fragments thereof derived from the viral agent for ET-NANB. Methods for preparation of ET-NANB viral proteins include isolating ET-NANB genomic sequences which are then cloned and expressed in a host cell. The resultant recombinant viral proteins find use as diagnostic agents and as vaccines. The genomic sequences and fragments thereof find use in preparing ET-NANB viral proteins and as probes for virus detection.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows vector constructions and manipulations used in obtaining and sequencing cloned ET-NANB fragment; and

Figures 2A-2B are representations of Southern blots in which a radiolabeled ET-NANB probe was hybridized with amplified cDNA fragments prepared from RNA isolated from infected (I) and non-infected (N) bile sources (2A), and from infected (I) and non-infected (N) stool-sample sources (2B).

DESCRIPTION OF SPECIFIC EMBODIMENTS

Novel compositions comprising generic sequences and fragments thereof derived from the viral agent for ET-NANB are provided, together with recombinant viral proteins produced using the genomic sequences and methods of using these compositions.

The genome of the ET-NANB viral agent is identified as containing a region which is homologous to the 1.33 kb DNA EcoRI insert present in plasmid pTZ-KF1 (ET1.1) carried in *E. coli* strain BB4 and having ATCC deposit no. 67717. Initial studies sequenced the two terminal regions of the insert and an intermediate region. The 5'-end region of this insert contains the sequence:

1 42
GAT GGA AGG CAC TAA TCT GGC AAG ACC TGT CCC TGT TGC AGC

43 84
TGT TCT ACC ACC CTG CCC CGA GCT CGA ACA GGG CCT TCT CTA

85 126
CCT GCC CCA GGA GCT CAC ACA CCC TGT GAT AGT GTC GTA ACA

127 168
TTT GAA TTA ACA GAC ATT GTG CAC TGC CGC ATG GCC GCC CCG

169 210
AGC CAG CGC AAG GCC GTG CTG TCC ACA CTC GTG GGC CGC TAC

211
GGC.

An intermediate region has the sequence:

691 731
CTA GAG TGT GCT ATT ATG GAG GAG TGT GGG ATG CCG CAG TGG

733 774
CTC ATC CGC CTG TAT CAC CTT ATA AGG TCT GCG TGG ATC TTG

775 816
CAG GCC CCG AAG GAG TCT CTG CGA GGG TTT TGG AAG AAA CAC

817 858
TCC GGT GAG CCC GGC ACT CTT CTA TGG AAT ACT GTC TGG AAT

859 900
ATG GCC GTT ATT ACC CAC TGT TAT GAC TTC CGC GAT TTT CAG

901 942
GTG GCT GCC TTT AAA GGT GAT GAT TCG ATA GTG CTT TGC AGT

943 984
GAG TAT CGT CAG AGT CCA GGA GCT GCT GTC CTG ATC GCC GGC

985 1026
TGT GGC TTG AAG TTG AAG GTA GAT TTC CGC CCG ATC GGT TTG

1027
TAT.

The 3'-end region contains the sequence:

1191 1232
TGA GTA GAG GAT GTT GTT TCC CGT GTT TAT GGG GTT TCC CCT

1233 1274
GGA CTC GTT CAT AAC CTG ATT GGC ATG CTA CAG GCT GTT GCT

1275 1316
GAT GGC AAG GCA CAT TTC ACT GAG TCA GTA AAA CCA GTG CTC

1317 1327
GAC CGG AAT TC.

Additional work has provided the entire sequence, in both directions, as set forth below. The sequence of both strands is provided, since it is not known in which strand the gene is located. However, the sequence in one direction has been designated as the "forward" sequence because of statistical similarities to known proteins. This sequence is set forth below along with the three possible translation sequences. There is one long open reading frame that starts at nucleotide 145 with an isoleucine and extends to the end of the sequence. The two other reading frames have many termination codons. Standard abbreviations for nucleotides and amino acids are used here and elsewhere in this specification.

Forward Sequence

```
R  I  P  P  T  D  G  R  H  Z  S  G  K  T  C  P  C  C  S  C
 E  F  R  Q  L  M  E  G  T  N  L  A  R  P  V  P  V  A  A  V
  N  S  A  N  Z  W  K  A  L  I  W  Q  D  L  S  L  L  Q  L  F
*          *          *          *          *          *
1          11         21         31         41         51
CGAATTCCGCCAACTGATGGAAGGCACTAATCTGGCAAGACCTGTCCCTGTTGCAGCTGT
```

```
 S  T  T  L  P  R  A  R  T  G  P  S  L  P  A  P  G  A  H  H
  L  P  P  C  P  E  L  E  Q  G  L  L  Y  L  P  Q  E  L  T  T
   Y  H  P  A  P  S  S  N  R  A  F  S  T  C  P  R  S  S  P  P
*         *         *         *         *         *
61        71        81        91        101       111
TCTACCACCCTGCCCCGAGCTCGAACAGGGCCTTCTCTACCTGCCCCAGGAGCTCACCAC

 L  Z  Z  C  R  N  I  Z  I  N  R  H  C  A  L  P  H  G  R  P
  C  D  S  V  V  T  F  E  L  T  D  I  V  H  C  R  M  A  A  P
   V  I  V  S  Z  H  L  N  Z  Q  T  L  C  T  A  A  W  P  P  R
*         *         *         *         *         *
121       131       141       151       161       171
CTGTGATAGTGTCGTAACATTTGAATTAACAGACATTGTGCACTGCCGCATGGCCGCCCC

 E  P  A  Q  G  R  A  V  H  T  R  G  P  L  R  R  R  T  K  L
  S  Q  R  K  A  V  L  S  T  L  V  G  R  Y  G  V  A  Q  S  S
   A  S  A  R  P  C  C  P  H  S  W  A  A  T  A  S  H  K  A  L
*         *         *         *         *         *
181       191       201       211       221       231
GAGCCAGCGCAAGGCCGTGCTGTCCACACTCGTGGGCCGCTACGGCGTCGCACAAAGCTC

 Y  N  A  S  H  S  D  V  R  D  S  L  A  R  F  I  P  A  I  G
  T  M  L  P  T  L  M  F  A  T  L  S  P  V  L  S  R  P  L  A
   Q  C  F  P  L  Z  C  S  R  L  S  R  P  F  Y  P  G  H  W  P
*         *         *         *         *         *
241       251       261       271       281       291
TACAATGCTTCCCACTCTGATGTTCGCGACTCTCTCGCCCGTTTTATCCCGGCCATTGGC

 P  V  Q  V  T  T  C  E  L  Y  E  L  V  E  A  M  V  E  K  G
  P  Y  R  L  Q  L  V  N  C  T  S  Z  W  R  P  W  S  R  R  A
   R  T  G  Y  N  L  Z  I  V  R  A  S  G  G  H  G  R  E  G  P
*         *         *         *         *         *
301       311       321       331       341       351
CCCGTACAGGTTACAACTTGTGAATTGTACGAGCTAGTGGAGGCCATGGTCGAGAAGGGC
```

```
Q D G S A V L E L D L C N R D V S R I T
 R M A P P S L S L I F A T V T C P G S P
  G W L R R P Z A Z S L Q P Z R V Q D H L
*         *         *         *         *         *
361       371       381       391       401       411
CAGGATGGCTCCGCCGTCCTTGAGCTTGATCTTTGCAACCGTGACGTGTCCAGGATCACC

F F Q K D C N K F T T G E T I A H G K V
 S S R K I V T S S P Q V R P L P M V K W
  L P E R L Z Q V H H R Z D H C P W Z S G
*         *         *         *         *         *
421       431       441       451       461       471
TTCTTCCAGAAAGATTGTAACAAGTTCACCACAGGTGAGACCATTGCCCATGGTAAAGTG

G Q G I S A W S K T F C A L F G P W F R
 A R A S R P G A R P S A P S L A L G S A
  P G H L G L E Q D L L R P L W P L V P R
*         *         *         *         *         *
481       491       501       511       521       531
GGCCAGGGCATCTCGGCCTGGAGCAAGACCTTCTGCGCCCTCTTTGGCCCTTGGTTCCGC

A I E K A I L A L L P Q G V F Y G D A F
 L L R R L P W P C S L R V C F T V M P L
  Y Z E G Y S G P A P S G C V L R Z C L Z
*         *         *         *         *         *
541       551       561       571       581       591
GCTATTGAGAAGGCTATTCTGGCCCTGCTCCCTCAGGGTGTGTTTTACGGTGATGCCTTT

D D T V F S A A V A A A K A S M V F E N
 M T P S S R R L W P Q Q R H P W C L R M
  Z H R L L G G C G R S K G I H G V Z E Z
*         *         *         *         *         *
601       611       621       631       641       651
GATGACACCGTCTTCTCGGCGGCTGTGGCCGCAGCAAAGGCATCCATGGTGTTTGAGAAT
```

```
 D  F  S  E  F  D  S  T  Q  N  N  F  S  L  G  L  E  C  A  I
  T  F  L  S  L  T  P  P  R  I  T  F  L  W  V  Z  S  V  L  L
   L  F  Z  V  Z  L  H  P  E  Z  L  F  S  G  S  R  V  C  Y  Y
*           *           *           *           *           *
661         671         681         691         701         711
GACTTTTCTGAGTTTGACTCCACCCAGAATAACTTTTCTCTGGGTCTAGAGTGTGCTATT

 M  E  E  C  G  M  P  Q  W  L  I  R  L  Y  H  L  I  R  S  A
  W  R  S  V  G  C  R  S  G  S  S  A  C  I  T  L  Z  G  L  R
   G  G  V  W  D  A  A  V  A  H  P  P  V  S  P  Y  K  V  C  V
*           *           *           *           *           *
721         731         741         751         761         771
ATGGAGGAGTGTGGGATGCCGCAGTGGCTCATCCGCCTGTATCACCTTATAAGGTCTGCG

 W  I  L  Q  A  P  K  E  S  L  R  G  F  W  K  K  H  S  G  E
  G  S  C  R  P  R  R  S  L  C  E  G  F  G  R  N  T  P  V  S
   D  L  A  G  P  E  G  V  S  A  R  V  L  E  E  T  L  R  Z  A
*           *           *           *           *           *
781         791         801         811         821         831
TGGATCTTGCAGGCCCCGAAGGAGTCTCTGCGAGGGTTTTGGAAGAAACACTCCGGTGAG

 P  G  T  L  L  W  N  T  V  W  N  M  A  V  I  T  H  C  Y  D
  P  A  L  F  Y  G  I  L  S  G  I  W  P  L  L  P  T  V  M  T
   R  H  S  S  M  E  Y  C  L  E  Y  G  R  Y  Y  P  L  L  Z  L
*           *           *           *           *           *
841         851         861         871         881         891
CCCGGCACTCTTCTATGGAATACTGTCTGGAATATGGCCGTTATTACCCACTGTTATGAC

 F  R  D  F  Q  V  A  A  F  K  G  D  D  S  I  V  L  C  S  E
  S  A  I  F  R  W  L  P  L  K  V  M  I  R  Z  C  F  A  V  S
   P  R  F  S  G  G  C  L  Z  R  Z  Z  F  D  S  A  L  Q  Z  V
*           *           *           *           *           *
901         911         921         931         941         951
TTCCGCGATTTTCAGGTGGCTGCCTTTAAAGGTGATGATTCGATAGTGCTTTGCAGTGAG
```

```
Y  R  Q  S  P  G  A  A  V  L  I  A  G  C  G  L  K  L  K  V
 I  V  R  V  Q  E  L  L  S  Z  S  P  A  V  A  Z  S  Z  R  Z
  S  S  E  S  R  S  C  C  P  D  R  R  L  W  L  E  V  E  G  R
*         *         *         *         *         *
961       971       981       991       1001      1011
TATCGTCAGAGTCCAGGAGCTGCTGTCCTGATCGCCGGCTGTGGCTTGAAGTTGAAGGTA

D  F  R  P  I  G  L  Y  A  G  V  V  V  A  P  G  L  G  A  L
 I  S  A  R  S  V  C  M  Q  V  L  W  W  P  P  A  L  A  R  S
  F  P  P  D  R  F  V  C  R  C  C  G  G  P  R  P  W  R  A  P
*         *         *         *         *         *
1021      1031      1041      1051      1061      1071
GATTTCCGCCCGATCGGTTTGTATGCAGGTGTTGTGGTGGCCCCCGGCCTTGGCGCGCTC

P  D  V  V  R  F  A  G  R  L  T  E  K  N  W  G  P  G  P  E
 L  M  L  C  A  S  P  A  G  L  P  R  R  I  G  A  L  A  L  S
  Z  C  C  A  L  R  R  P  A  Y  R  E  E  L  G  P  W  P  Z  A
*         *         *         *         *         *
1081      1091      1101      1111      1121      1131
CCTGATGTTGTGCGCTTCGCCGGCCGGCTTACCGAGAAGAATTGGGGCCCTGGCCCTGAG

R  A  E  Q  L  R  L  A  V  S  D  F  L  R  K  L  T  N  V  A
 G  R  S  S  S  A  S  L  L  V  I  S  S  A  S  S  R  M  Z  L
  G  G  A  A  P  P  R  C  Z  Z  F  P  P  Q  A  H  E  C  S  S
*         *         *         *         *         *
1141      1151      1161      1171      1181      1191
CGGGGGGAGCAGCTCCGCCTCGCTGTTAGTGATTTCCTCCGCAAGCTCACGAATGTAGCT

Q  M  C  V  D  V  V  S  R  V  Y  G  V  S  P  G  L  V  H  N
 R  C  V  W  M  L  F  P  V  F  M  G  F  P  L  D  S  F  I  T
  D  V  C  G  C  C  F  P  C  L  W  G  F  P  W  T  R  S  Z  P
*         *         *         *         *         *
1201      1211      1221      1231      1241      1251
CAGATGTGTGTGGATGTTGTTTCCCGTGTTTATGGGGTTTCCCCTGGACTCGTTCATAAC
```

```
 L   I   G   M   L   Q   A   V   A   D   G   K   A   H   F   T   E   S   V   K
   Z   L   A   C   Y   R   L   L   L   M   A   R   H   I   S   L   S   Q   Z   N
     D   W   H   A   T   G   C   C   Z   W   Q   G   T   F   H   Z   V   S   K   T
*                *                *                *                *                *
1261             1271             1281             1291             1301             1311
CTGATTGGCATGCTACAGGCTGTTGCTGATGGCAAGGCACATTTCACTGAGTCAGTAAAA

 P   V   L   D   R   N   S   S
   Q   C   S   T   G   I   R
     S   A   R   P   E   F   E
*                *                *
1321             1331             1341
CCAGTGCTCGACCGGAATTCGAGC
```

The complimentary strand, referred to here as the "reverse sequence," is set forth below in the same manner as the forward sequence set forth above. Several open reading frames, shorter than the long open reading frame found in the forward sequence, can be seen in this reverse sequence.

Reverse Sequence

```
 A   R   I   P   V   E   H   W   F   Y   Z   L   S   E   M   C   L   A   I   S
   L   E   F   R   S   S   T   G   F   T   D   S   V   K   C   A   L   P   S   A
     S   N   S   G   R   A   L   V   L   L   T   Q   Z   N   V   P   C   H   Q   Q
*                *                *                *                *                *
1                11               21               31               41               51
GCTCGAATTCCGGTCGAGCACTGGTTTTACTGACTCAGTGAAATGTGCCTTGCCATCAGC

 N   S   L   Z   H   A   N   Q   V   M   N   E   S   R   G   N   P   I   N   T
   T   A   C   S   M   P   I   R   L   Z   T   S   P   G   E   T   P   Z   T   R
     Q   P   V   A   C   Q   S   G   Y   E   R   V   Q   G   K   P   H   K   H   G
*                *                *                *                *                *
61               71               81               91               101              111
AACAGCCTGTAGCATGCCAATCAGGTTATGAACGAGTCCAGGGGAAACCCCATAAACACG
```

```
 G  N  N  I  H  T  H  L  S  Y  I  R  E  L  A  E  E  I  T  N
  E  T  T  S  T  H  I  Z  A  T  F  V  S  L  R  R  K  S  L  T
   K  Q  H  P  H  T  S  E  L  H  S  Z  A  C  G  G  N  H  Z  Q
*         *         *         *         *         *
121       131       141       151       161       171
GGAAACAACATCCACACACATCTGAGCTACATTCGTGAGCTTGCGGAGGAAATCACTAAC

 S  E  A  E  L  L  R  P  L  R  A  R  A  P  I  L  L  G  K  P
  A  R  R  S  C  S  A  R  S  G  P  G  P  Q  F  F  S  V  S  R
   R  G  G  A  A  P  P  A  Q  G  Q  G  P  N  S  S  R  Z  A  G
*         *         *         *         *         *
181       191       201       211       221       231
AGCGAGGCGGAGCTGCTCCGCCCGCTCAGGGCCAGGGCCCCAATTCTTCTCGGTAAGCCG

 A  G  E  A  H  N  I  R  E  R  A  K  A  G  G  H  H  N  T  C
  P  A  K  R  T  T  S  G  S  A  P  R  P  G  A  T  T  T  P  A
   R  R  S  A  Q  H  Q  G  A  R  Q  G  R  G  P  P  Q  H  L  H
*         *         *         *         *         *
241       251       261       271       281       291
GCCGGCGAAGCGCACAACATCAGGGAGCGCGCCAAGGCCGGGGGCCACCACAACACCTGC

 I  Q  T  D  R  A  E  I  Y  L  Q  L  Q  A  T  A  G  D  Q  D
  Y  K  P  I  G  R  K  S  T  F  N  F  K  P  Q  P  A  I  R  T
   T  N  R  S  G  G  N  L  P  S  T  S  S  H  S  R  R  S  G  Q
*         *         *         *         *         *
301       311       321       331       341       351
ATACAAACCGATCGGGCGGAAATCTACCTTCAACTTCAAGCCACAGCCGGCGATCAGGAC

 S  S  S  W  T  L  T  I  L  T  A  K  H  Y  R  I  I  T  F  K
  A  A  P  G  L  Z  R  Y  S  L  Q  S  T  I  E  S  S  P  L  K
   Q  L  L  D  S  D  D  T  H  C  K  A  L  S  N  H  H  L  Z  R
*         *         *         *         *         *
361       371       381       391       401       411
AGCAGCTCCTGGACTCTGACGATACTCACTGCAAAGCACTATCGAATCATCACCTTTAAA
```

```
G  S  H  L  K  I  A  E  V  I  T  V  G  N  N  G  H  I  P  D
 A  A  T  Z  K  S  R  K  S  Z  Q  W  V  I  T  A  I  F  Q  T
  Q  P  P  E  N  R  G  S  H  N  S  G  Z  Z  R  P  Y  S  R  Q
*         *         *         *         *         *
421       431       441       451       461       471
GGCAGCCACCTGAAAATCGCGGAAGTCATAACAGTGGGTAATAACGGCCATATTCCAGAC

S  I  P  Z  K  S  A  G  L  T  G  V  F  L  P  K  P  S  Q  R
 V  F  H  R  R  V  P  G  S  P  E  C  F  F  Q  N  P  R  R  D
  Y  S  I  E  E  C  R  A  H  R  S  V  S  S  K  T  L  A  E  T
*         *         *         *         *         *
481       491       501       511       521       531
AGTATTCCATAGAAGAGTGCCGGGCTCACCGGAGTGTTTCTTCCAAAACCCTCGCAGAGA

L  L  R  G  L  Q  D  P  R  R  P  Y  K  V  I  Q  A  D  E  P
 S  F  G  A  C  K  I  H  A  D  L  I  R  Z  Y  R  R  M  S  H
  P  S  G  P  A  R  S  T  Q  T  L  Z  G  D  T  G  G  Z  A  T
*         *         *         *         *         *
541       551       561       571       581       591
CTCCTTCGGGGCCTGCAAGATCCACGCAGACCTTATAAGGTGATACAGGCGGATGAGCCA

L  R  H  P  T  L  L  H  N  S  T  L  Z  T  Q  R  K  V  I  L
 C  G  I  P  H  S  S  I  I  A  H  S  R  P  R  E  K  L  F  W
  A  A  S  H  T  P  P  Z  Z  H  T  L  D  P  E  K  S  Y  S  G
*         *         *         *         *         *
601       611       621       631       641       651
CTGCGGCATCCCACACTCCTCCATAATAGCACACTCTAGACCCAGAGAAAAGTTATTCTG

G  G  V  K  L  R  K  V  I  L  K  H  H  G  C  L  C  C  G  H
 V  E  S  N  S  E  K  S  F  S  N  T  M  D  A  F  A  A  A  T
  W  S  Q  T  Q  K  S  H  S  Q  T  P  W  M  P  L  L  R  P  Q
*         *         *         *         *         *
661       671       681       691       701       711
GGTGGAGTCAAACTCAGAAAAGTCATTCTCAAACACCATGGATGCCTTTGCTGCGGCCAC
```

```
S  R  R  E  D  G  V  I  K  G  I  T  V  K  H  T  L  R  E  Q
 A  A  E  K  T  V  S  S  K  A  S  P  Z  N  T  P  Z  G  S  R
  P  P  R  R  R  C  H  Q  R  H  H  R  K  T  H  P  E  G  A  G
*         *         *         *         *         *
721       731       741       751       761       771
AGCCGCCGAGAAGACGGTGTCATCAAAGGCATCACCGTAAAACACACCCTGAGGGAGCAG

G  Q  N  S  L  L  N  S  A  E  P  R  A  K  E  G  A  E  G  L
 A  R  I  A  F  S  I  A  R  N  Q  G  P  K  R  A  Q  K  V  L
  P  E  Z  P  S  Q  Z  R  G  T  K  G  Q  R  G  R  R  R  S  C
*         *         *         *         *         *
781       791       801       811       821       831
GGCCAGAATAGCCTTCTCAATAGCGCGGAACCAAGGGCCAAAGAGGGCGCAGAAGGTCTT

A  P  G  R  D  A  L  A  H  F  T  M  G  N  G  L  T  C  G  E
 L  Q  A  E  M  P  W  P  T  L  P  W  A  M  V  S  P  V  V  N
  S  R  P  R  C  P  G  P  L  Y  H  G  Q  W  S  H  L  W  Z  T
*         *         *         *         *         *
841       851       861       871       881       891
GCTCCAGGCCGAGATGCCCTGGCCCACTTTACCATGGGCAATGGTCTCACCTGTGGTGAA

L  V  T  I  F  L  E  E  G  D  P  G  H  V  T  V  A  K  I  K
 L  L  Q  S  F  W  K  K  V  I  L  D  T  S  R  L  Q  R  S  S
  C  Y  N  L  S  G  R  R  Z  S  W  T  R  H  G  C  K  D  Q  A
*         *         *         *         *         *
901       911       921       931       941       951
CTTGTTACAATCTTTCTGGAAGAAGGTGATCCTGGACACGTCACGGTTGCAAAGATCAAG

L  K  D  G  G  A  I  L  A  L  L  D  H  G  L  H  Z  L  V  Q
 S  R  T  A  E  P  S  W  P  F  S  T  M  A  S  T  S  S  Y  N
  Q  G  R  R  S  H  P  G  P  S  R  P  W  P  P  L  A  R  T  I
*         *         *         *         *         *
961       971       981       991       1001      1011
CTCAAGGACGGCGGAGCCATCCTGGCCCTTCTCGACCATGGCCTCCACTAGCTCGTACAA
```

```
 F  T  S  C  N  L  Y  G  A  N  G  R  D  K  T  G  E  R  V  A
  S  Q  V  V  T  C  T  G  P  M  A  G  I  K  R  A  R  E  S  R
   H  K  L  Z  P  V  R  G  Q  W  P  G  Z  N  G  R  E  S  R  E
*         *         *         *         *         *
1021      1031      1041      1051      1061      1071
TTCACAAGTTGTAACCTGTACGGGGCCAATGGCCGGGATAAAACGGGCGAGAGAGTCGCG

 N  I  R  V  G  S  I  V  E  L  C  A  T  P  Z  R  P  T  S  V
  T  S  E  W  E  A  L  Z  S  F  V  R  R  R  S  G  P  R  V  W
   H  Q  S  G  K  H  C  R  A  L  C  D  A  V  A  A  H  E  C  G
*         *         *         *         *         *
1081      1091      1101      1111      1121      1131
AACATCAGAGTGGGAAGCATTGTAGAGCTTTGTGCGACGCCGTAGCGGCCCACGAGTGTG

 D  S  T  A  L  R  W  L  G  A  A  M  R  Q  C  T  M  S  V  N
  T  A  R  P  C  A  G  S  G  R  P  C  G  S  A  Q  C  L  L  I
   Q  H  G  L  A  L  A  R  G  G  H  A  A  V  H  N  V  C  Z  F
*         *         *         *         *         *
1141      1151      1161      1171      1181      1191
GACAGCACGGCCTTGCGCTGGCTCGGGGCGGCCATGCGGCAGTGCACAATGTCTGTTAAT

 S  N  V  T  T  L  S  Q  V  V  S  S  W  G  R  Z  R  R  P  C
  Q  M  L  R  H  Y  H  R  W  Z  A  P  G  A  G  R  E  G  P  V
   K  C  Y  D  T  I  T  G  G  E  L  L  G  Q  V  E  K  A  L  F
*         *         *         *         *         *
1201      1211      1221      1231      1241      1251
TCAAATGTTACGACACTATCACAGGTGGTGAGCTCCTGGGGCAGGTAGAGAAGGCCCTGT

 S  S  S  G  Q  G  G  R  T  A  A  T  G  T  G  L  A  R  L  V
  R  A  R  G  R  V  V  E  Q  L  Q  Q  G  Q  V  L  P  D  Z  C
   E  L  G  A  G  W  Z  N  S  C  N  R  D  R  S  C  Q  I  S  A
*         *         *         *         *         *
1261      1271      1281      1291      1301      1311
TCGAGCTCGGGGCAGGGTGGTAGAACAGCTGCAACAGGGACAGGTCTTGCCAGATTAGTG
```

P S I S W R N S
L P S V G G I
F H Q L A E F
* * *
1321 1331 1341
CCTTCCATCAGTTGGCGGAATTCG

Identity of this sequence with sequences in etiologic agents has been confirmed by locating a corresponding sequence in a viral strain isolated in Burma. The Burmese isolate contains the following sequence of nucleotides.

Non-A Non-B ET: Burmese strain

```
  1 CGGTTGTTCA GTACCAGTTT ACTGCAGGTG TGCCTGGATC CGGCAAGTCC
 51 CGCTCTATCA CCCAAGCCGA TGTGGACGTT GTCGTGGTCC CGACGCGTGA
101 GTTGCGTAAT GCCTGGCGCC GTCGCGGCTT TGCTGCTTTT ACCCCGCATA
151 CTGCCGCCAG AGTCACCCAG GGGCGCCGGG TTGTCATTGA TGAGGCTCCA
201 TCCCTCCCCC CTCACCTGCT GCTGCTCCAC ATGCAGCGGG CCGCCACCGT
251 CCACCTTCTT GGCGCCCCGA ACCAGATCCC AGCCATCGAC TTTGAGCACG
301 CTGGGCTCGT CCCCGCCATC AGGCCCGACT TAGCCCCACC TCCTGGTGGC
351 ATGTTACCCA TCGCTGCCCT GCGGATGTAT GCGAGCTCAT CCGTGGTGCA
401 TACCCCATGA TCCAGACCAC TAGCCGGGTT CTCCGTTCGT TGTTCTGGGG
451 TGAGCCTGCC GTCGGGCAGA AACTAGTGTT CACCCAGGCG GCCAAGGCCG
501 CCAACCCCGG CTCAGTGACG GTCCACGAGG CGCAGGGCGC TACCTACACG
551 GAGACCACTA TTATTGCCAC AGCAGATGCC CGGGGCCTTA TTCAGTCGTC
601 TCGGGCTCAT GCCATTGTTG CTCTGACGCG CCACACTGAG AAGTGCGTCA
651 TCATTGACGC ACCAGGCCTG CTTCGCGAGG TGGGCATCTC CGATGCAATC
701 GTTAATAACT TTTTCCTCGC TGGTGGCGAA ATTGGTCACC AGCGCCCATC
751 AGTTATTCCC CGTGGCAACC CTGACGCCAA TGTTGACACC CTGGCTGCCT
801 TCCCGCCGTC TTGCCAGATT AGTGCCTTCC ATCAGTTGGC TGAGGAGCTT
851 GGCCACAGAC CTGTCCCTGT TGCAGCTGTT CTACCACCCT GCCCCGAGCT
901 CGAACAGGGC CTTCTCTACC TGCCCCAGGA GCTCACCACC TGTGATAGTG
951 TCGTAACATT TGAATTAACA GACATTGTGC ACTGCCGCAT GGCCGCCCCG
```

1001 AGCCAGCGCA AGGCCGTGCT GTCCACACTC GTGGGCCGCT ACGGCGTCGC
1051 ACAAAGCTCT ACAATGCTTC CCACTCTGAT GTTCGCGACT CTCTCGCCCG
1101 TTTTATCCCG GCCATTGGCC CCGTACAGGT TACAACTTGT GAATTGTACG
1151 AGCTAGTGGA GGCCATGGTC GAGAAGGGCC AGGATGGCTC CGCCGTCCTT
1201 GAGCTTGATC TTTGCAACCG TGACGTGTCC AGGATCACCT TCTTCCAGAA
1251 AGATTGTAAC AAGTTCACCA CAGGTGAGAC CATTGCCCAT GGTAAAGTGG
1301 GCCAGGGCAT CTCGGCCTGG AGCAAGACCT TCTGCGCCCT CTTTGGCCCT
1351 TGGTTCCGCG CTATTGAGAA GGCTATTCTG GCCCTGCTCC CTCAGGGTGT
1401 GTTTTACGGT GATGCCTTTG ATGACACCGT CTTCTCGGCG GCTGTGGCCG
1451 CAGCAAAGGC ATCCATGGTG TTTGAGAATG ACTTTTCTGA GTTTGACTCC
1501 ACCCAGAATA ACTTTTCTCT GGGTCTAGAG TGTGCTATTA TGGAGGAGTG
1551 TGGGATGCCG CAGTGGCTCA TCCGCCTGTA TCACCTTATA AGGTCTGCGT
1601 GGATCTTGCA GGCCCCGAAG GAGTCTCTGC GAGGGTTTTG GAAGAAACAC
1651 TCCGGTGAGC CCGGCACTCT TCTATGGAAT ACTGTCTGGA ATATGGCCGT
1701 TATTACCCAC TGTTATGACT TCCGCGATTT TCAGGTGGCT GCCTTTAAAG
1751 GTGATGATTC GATAGTGCTT TGCAGTGAGT ATCGTCAGAG TCCAGGAGCT
1801 GCTGTCCTGA TCGCCGGCTG TGGCTTGAAG TTGAAGGTAG ATTTCCGCCC
1851 GATCGGTTTG TATGCAGGTG TTGTGGTGGC CCCCGGCCTT GGCGCGCTCC
1901 CTGATGTTGT GCGCTTCGCC GGCCGGCTTA CCGAGAAGAA TTGGGGCCCT
1951 GGCCCTGAGC GGGCGGAGCA GCTCCGCCTC GCTGTTAGTG ATTTCCTCCG
2001 CAAGCTCACG AATGTAGCTC AGATGTGTGT GGATGTTGTT TCCCGTGTTT
2051 ATGGGGTTTC CCCTGGACTC GTTCATAACC TGATTGGCAT GCTACAGGCT
2101 GTTGCTGATG GCAAGGCACA TTTCACTGAG TCAGTAAAAC CAGTGCTCGA
2151 CTTGACAAAT TCAATCTTGT GTCGGGTGGA ATGAATAACA TGTCTTTTGC
2201 TGCGCCCATG GGTTCGCGAC CATGCGCCCT CGGCCTATTT TGTTGCTGCT
2251 CCTCATGTTT TTGCCTATGC TGCCCGCGCC ACCGCCCGGT CAGCCGTCTG
2301 GCCGCCGTCG TGGGCGGCGC AGCGGCGGTT CCGGCGGTGG TTTCTGGGGT
2351 GACCGGGTTG ATTCTCAGCC CTTCGCAATC CCCTATATTC ATCCAACCAA
2401 CCCCTTCGCC CCCGATGTCA CCGCTGCGGC CGGGGCTGGA CCTCGTGTTC
2451 GCCAACCCGC CCGACCACTC GCGTCCGCTT GGCGTGACCA GGCCCAGCGC
2501 CCCGCCGTTG CCTCACGTCG TAGACCTACC ACAGCTGGGG CCGCGCCGCT
2551 AACCGCGGTC GCTCCGGCCC CG

The ability of the methods described herein to isolate and identify genetic material from other NANB hepatitis strains has been confirmed by identifying genetic material from an isolate obtained in Mexico.

The sequence of this isolate was about 75% identical to the ET1.1 sequence set forth above. The sequence was identified by hybridization using the conditions set forth in Section II.B below. A partial cDNA sequence consisting of 1661 nucleotides is set forth below.

Non-A Non-B ET: Mexican Strain

```
   1 GCGGCCGCTC GGTGGGGTTC CACCTCTCTT CCCATTCCGA ACAGAGAAGT
  51 TAAGCAAGGA ATTAATTCGC GGCCGCTCGT GTTGCGTGAG GTGGGTATCT
 101 CAGATGCCAT TGTTAATAAT TTCTTCCTTT CGGGTGGCGA GGTTGGTCAC
 151 CAGAGACCAT CGGTCATTCC GCGAGCAACC CTGACCGCAA TGTTGACGTG
 201 CTTGCGGCGT TTCCACCTTC ATGCCAAATA AGCGCCTTCC ATCAGCTTGC
 250 TGAGGAGCTG GGCCACCGGC GGCGCCGGTG CTGTGCTACC TCCCTGCCCT
 301 GAGCTTGAGC AGGGCCTTCT CTATCTGCCA CAGGAGCTAG CCTCCTGTGA
 350 CAGTGTTGTG ACATTTGAGC TAACTGACAT TGTGCACTGC CGCATGGCGG
 401 CCCCTAGCCA AAGGAAAGCT GTTTTGTCCA CGCTGGTAGG CCGGTATGGC
 451 AGACGCACAA GGCTTTATGA TGCGGGTCAC ACCGATGTCC GCGCCTCCCT
 501 TGCGCGCTTT ATTCCCACTC TCGGGCGGGT TACTGCCACC ACCTGTGAAC
 551 TCTTTGAGCT TGTAGAGGCG ATGGTGGAGA AGGGCCAAGA CGGTTCAGCC
 601 GTCCTCGAGT TGGATTTGTG CAGCCGAGAT GTCTCCCGCA TAACCTTTTT
 651 CCAGAAGGAT TGTAACAAGT TCACGACCGG CGAGACAATT GCGCATGGCA
 701 AAGTCGGTCA GGGTATCTCC GCCTGGAGTA AGACCTTTTG TGCCCTGTTT
 751 GGCCCCTGGT TCCGTTGCGA TTGAGAAGGC TATTCTATCC CTTTTACCAC
 801 AAGCTGTGTT CTACGGGGAT GCTTATGACG ACTCAGTATT CTCTGCTGCC
 851 GTGGCTGGCG CCAGCCATGC CATGGTGTTT GAAAATGATT TTTCTGAGTT
 901 TGACTCGACT CAGAATAACT TTTCCCTAGG TCTTGAGTGC GCCATTATGG
 951 AAGAGTGTGG TATGCCCCAG TGGCTTGTCA GGTTGTACCA TGCCGTCCGG
1001 TCGGCGTGGA TCCTGCAGGC CCCAAAAGAG TCTTTGAGAG GGTTCTGGAA
1051 GAAGCATTCT GGTGAGCCGG GCACGTTGCT CTGGAATACG GTGTGGAACA
1101 TGGCAATCAT TGCCCATTGC TATGAGTTCC GGGACCTCCA GGTTGCCGCC
1151 TTCAAGGGCG ACGACTCGGT CGTCCTCTGT AGTGAATACC GCCAGAGCCC
1201 AGGCGCCGGT TCGCTTATAG CAGGCTGTGG TTTGAAGCCG AAGGCTGACT
1251 TCCGGCCGAT TGGGCTGTAT GCCGGGGTTG TCGTCGCCCC GGGGCTCGGG
1301 GCCCTACCCG ATGTCGTTCG ATTCGCCGGA CGGCTTTCGG AGAAGAACTG
1351 GGGGCCTGAT CCGGAGCGGG CAGAGCAGCT CCGCCTCGCC GTGCAGGATT
1401 TCCTCCGTAG GTTAACGAAT GTGGCCCAGA TTTGTGTTGA GGTGGTGTCT
```

1451 AGAGTTTACG GGGTTTCCCC GGGTCTGGTT CATAACCTGA TAGGCATGCT
1501 CCAGACTATT GGTGATGGTA AGGCGCATTT TACAGAGTCT GTTAAGCCTA
1551 TACTTGACCT TACACACTCA ATTATGCACC GGTCTGAATG AATAACATGT
1601 GGTTTCCTGC GCCCATGGGT TCGCCACCAT GCGCCCTAGG CCTCTTTTGC
1651 CGAGCGGCCG C

When comparing the Burmese and Mexican strains, 76.7% identity is seen in a 1372 nucleotide overlap beginning at nucleotide 13 of the Mexican strain and nucleotide 331 of the Burmese strain.

It will be recognized by one skilled in the art of molecular genetics that each of the last two sequences teaches a corresponding complementary DNA sequence as well as RNA sequences corresponding to both the principal sequence shown and the complementary DNA sequence. Additionally, open reading frames encoding peptides are present, and expressible peptides are disclosed by the nucleotide sequences without setting forth the amino acid sequences explicitly, in the same manner as if the amino acid sequences were explicitly set forth as in the ET1.1 sequence above.

DETAILED DESCRIPTION OF THE INVENTION

I. Definitions

The terms defined below have the following meaning herein:

1. "Enterically transmitted non-A/non-B (ET-NANB) hepatitis viral agent" means a virus, virus type, or virus class which (i) causes water-borne, infectious hepatitis, (ii) is transmissible in cynomolgus monkeys, (iii) is serologically distinct from hepatitis A virus (HAV) and hepatitis B virus (HAB), and (iv) includes a genomic region which is homologous to the 1.33 kb cDNA

insert in plasmid pTZ-KF1(ET1.1) carried in E. coli strain BB4 identified by ATCC deposit number 67717.

2. Two nucleic acid fragments are "homologous" if they are capable of hybridizing to one another under hybridization conditions in which hybridized strands contain at most about 25-30% basepair mismatches. In general, two single-strand nucleic acid species will be homologous if they hybridize under the conditions described in Maniatis et al., op. cit., pp. 320-323, but using the following wash conditions: 2 x SCC, 0.1% SDS, room temperature - twice, 30 minutes each; then 2 x SCC, 0.1% SDS, 50°C - once, 30 minutes; then 2 x SCC, room temperature - twice, 10 minutes each. More preferably, homologous nucleic acid strands contain 15-25% basepair mismatches, even more preferably 5-15% basepair mismatches. These degrees of homology can be selected by using more stringent wash conditions for identification of clones from gene libraries (or other sources of genetic material), as is well known in the art.

3. A DNA fragment is "derived from" an ET-NANB viral agent if it has the same or substantially the same basepair sequence as a region of the viral agent genome.

4. A protein is "derived from" an ET-NANB viral agent if it is encoded by an open reading frame of a DNA or RNA fragment derived from an ET-NANB viral agent.

II. Obtaining Cloned ET-NANB Fragments

According to one aspect of the invention, it has been found that a virus-specific DNA clone can be produced by (a) isolating RNA from the bile of a cynomolgus monkey having a known ET-NANB infection, (b) cloning the cDNA fragments to form a fragment library, and (c) screening the library by differential hybridization to radiolabeled cDNAs from infected and non-infected bile sources.

A. cDNA Fragment Mixture

ET-NANB infection in cynomolgus monkeys is initiated by inoculating the animals intravenously with a 10% w/v suspension from human case stools positive for 27-34 nm ET-NANB particles (mean diameter 32 nm). An infected animal is monitored for elevated levels of alanine aminotransferase, indicating hepatitis infection. ET-NANB infection is confirmed by immunospecific binding of seropositive antibodies to virus-like particles (VLPs), according to published methods (Gravelle). Briefly, a stool (or bile) specimen taken from the infected animal 3-4 weeks after infection is diluted 1:10 with phosphate-buffered saline, and the 10% suspension is clarified by low-speed centrifugation and filtration successively through 1.2 and 0.45 micron filters. The material may be further purified by pelleting through a 30% sucrose cushion (Bradley). The resulting preparation of VLPs is mixed with diluted serum from human patients with known ET-NANB infection. After incubation overnight, the mixture is centrifuged overnight to pellet immune aggregates, and these are stained and examined by electron microscopy for antibody binding to the VLPs.

ET-NANB infection can also be confirmed by seroconversion to VLP-positive serum. Here the serum of the infected animal is mixed as above with 27-34 nm VLPs isolated form the stool specimens of infected human cases and examined by immune electron microscopy for antibody binding to the VLPs.

Bile can be collected from ET-NANB positive animals by either cannulating the bile duct and collecting the bile fluid or by draining the bile duct during necropsy. Total RNA is extracted from the bile by hot phenol extraction, as outlined in Example 1A. The RNA fragments are used to synthesize corresponding duplex cDNA fragments by random priming, also as

referenced in Example 1A. The cDNA fragments may be fractionated by gel electrophoresis or density gradient centrifugation to obtain a desired size class of fragments, e.g., 500-4,000 basepair fragments.

Although alternative sources of viral material, such as VLPs obtained from stool samples (as described in Example 4), may be used for producing a cDNA fraction, the bile source is preferred. According to one aspect of the invention, it has been found that bile from ET-NANB-infected monkeys shows a greater number of intact viral particles than material obtained from stool samples, as evidenced by immune electron microscopy. Bile obtained from an ET-NANB infected human or cynomolgus monkey, for use as a source of ET-NANB viral protein or genomic material, or intact virus, forms part of the present invention.

B. cDNA Library and Screening

The cDNA fragments from above are cloned into a suitable cloning vector to form a cDNA library. This may be done by equipping blunt-ended fragments with a suitable end linker, such as an EcoRI sequence, and inserting the fragments into a suitable insertion site of a cloning vector, such as at a unique EcoRI site. After initial cloning, the library may be recloned, if desired, to increase the percentage of vectors containing a fragment insert. The library construction described in Example 1B is illustrative. Here cDNA fragments were blunt-ended, equipped with EcoRI ends, and inserted into the EcoRI site of the lambda phage vector gt10. The library phage, which showed less than 5% fragment inserts, was isolated, and the fragment inserts recloned into the lambda gt10 vector, yielding more than 95% insert-containing phage.

The cDNA library is screened for sequences specific for ET-NANB by differential hybridization to cDNA probes derived from infected and non-infected

sources. cDNA fragments from infected and non-infected source bile or stool viral isolates can be prepared as above. Radiolabeling the fragments is by random labeling, nick translation, or end labeling, according to conventional methods (Maniatis, p. 109). The cDNA library from above is screened by transfer to duplicate nitrocellulose filters, and hybridization with both infected-source and non-infected-source (control) radiolabeled probes, as detailed in Example 2. In order to recover sequences that hybridize at the preferred outer limit of 25-30% basepair mismatches, clones can be selected if they hybridize under the conditions described in Maniatis et al., op. cit., pp. 320-323, but using the following wash conditions: 2 x SCC, 0.1% SDS, room temperature - twice, 30 minutes each; then 2 x SCC, 0.1% SDS, 50°C - once, 30 minutes; then 2 x SCC, room temperature - twice, 10 minutes each. These conditions allowed identification of the Mexican isolate discussed above using the ET1.1 sequence as a probe. Plaques which show selective hybridization to the infected-source probes are preferably replated at low plating density and rescreened as above, to isolate single clones which are specific for ET-NANB sequences. As indicated in Example 2, sixteen clones which hybridized specifically with infected-source probes were indentified by these procedures. One of the clones, designated lambda gt10-1.1, contained a 1.33 kilobase fragment insert.

C. ET-NANB Sequences

The basepair sequence of cloned regions of the ET-NANB fragments from Part B are determined by standard sequencing methods. In one illustrative method, described in Example 3, the fragment insert from the selected cloning vector is excised, isolated by gel electrophoresis, and inserted into a cloning vector whose basepair sequence on either side of the

insertion site is known. The particular vector employed in Example 3 is a pTZ-KF1 vector shown at the left in Figure 1. The ET-NANB fragment from the gt10-1.1 phage was inserted at the unique EcoRI site of the pTZ-KF1 plasmid. Recombinants carrying the desired insert were identified by hybridization with the isolated 1.33 kilobase fragment, as described in Example 3. One selected plasmid, identified as pTZ-KF1 (ET1.1), gave the expected 1.33 kb fragment after vector digestion with EcoRI. *E. coli* strain BB4 infected with the pTZ-KF1(ET1.1) plasmid has been deposited with the American Type Culture Collection, Rockville, MD, and is identified by ATCC deposit nubmer 67717.

The pTZ-KF1(ET1.1) plasmid is illustrated at the bottom in Figure 1. The fragment insert has 5' and 3' end regions denoted at A and C, respectively, and an intermediate region, denoted at B. The sequences in these regions were determined by standard dideoxy sequencing and are set forth above. The three short sequences (A, B, and C) are from the same insert strand. As will be seen in Example 3, the B-region sequence was actually determined from the opposite strand, so that the B-region sequence shown above represent the complement of the sequence in the sequenced strand. The base numbers of the partial sequences are approximate.

Later work in the laboratory of the inventors identified the full sequence, also set forth above. Fragments of this total sequence can readily be prepared using restriction endonucleases. Computer analysis of both the forward and reverse sequence has identified a number of cleavage sites. The specific cleavage sites are summarized (for the forward direction) in the following tables.

Pattern identifier (^ = cleavage site)	Pattern matched	Base number matched in "forward" strand
CC*WGG(BstNI)	CCWGG	106, 360, 410, 483, 497, 973, 1129, 1243
CC*SGG(NciI)	CCSGG	288, 841, 1063
GAAGANNNNNNNN*(MboII)	GAAGA	822, 1116
TCTTC(<-7-MboII)	TCTTC	422, 611, 849
GCGTC(<-10-HgaI)	GCGTC	225
GCATCNNNNN*(SfaNI)	GCATC	488, 640
GCAGCNNNNNNNN*(BbvI)	GCAGC	53, 631, 1149
GCTGC(<-12-BbvI)	GCTGC	919, 979
GGATGNNNNNNNNN*(FokI)	GGATG	363, 734, 1212
CATCC(<-13-FokI)	CATCC	641, 750
GGTGANNNNNNNN*(HphI)	GGTGA	454, 589, 835, 931
TCACC(<-7-HphI)	TCACC	114, 416, 446, 762
GP*CGYC(AhaII)	GPCGYC	224
GDGCH*C(BspI1286)	GDGCHC	77, 110, 158, 838, 1125, 1324
GPGCY*C(BanII)	GPGCYC	77, 110, 838, 1125
C*YCGPG(AvaI)	CYCGPG	74, 178
Y*GGCCP(EaeI)	YGGCCP	171, 290, 626, 875, 1101
GWGCW*C(GgiAI)	GWGCWC	77, 110, 158, 1324
C*CTTGG(StyI)	CCTTGG	529, 1068
P*GATCY(XhoII)	PGATCY	782
CAG*CTG(PvuII)	CAGCTG	54
C*CATGG(NcoI)	CCATGG	344, 468, 644
CGAT*CG(PvuI)	CGATCG	1031
C*GGCCG(EagI)	CGGCCG	1101
G*AATTC(EcoR1)	GAATTC	2, 1335
GAGCT*C(SacI)	GAGCTC	77, 110
GCATG*C(SphI)	GCATGC	1268
GCC*GGC(NaeI)	GCCGGC	994, 1099, 1103
G*CGCGC(BssHII)	GCGCGC	1073
GGGCC*C(ApaI)	GGGCCC	1125

TCG*CGA(NruI)	TCGCGA	264		
T*CTAGA(XbaI)	TCTAGA	705		
TTT*AAA(DraI)	TTTAAA	925		
G*TGCAC(ApaLI)	GTGCAC	158		
ACCTGCNNNN*(BpsMI)	ACCTGC	99		
GCAGGT(<-8-BspMI)	GCAGGT	1045		
GACN*NNGTC(Tth111I)	GACNNNGTC	604		
CCANNNN*NTGG(PflMI)	CCANNNNNTGG	10		
CC*TNAGG(MstII)	CCTNAGG	571		
GCCNNNN*NGGC(BglI)	GCCNNNNNGGC	216,	359,	738
CCANNNNN*NTGG(BstXI)	CCANNNNNNTGG	204		

III. <u>ET-NANB Fragments</u>

According to another aspect, the invention includes ET-NANB-specific fragments or probes which hybridize with ET-NANB genomic sequences or cDNA fragments derived therefrom. The fragments may include full-length cDNA fragments such as described in Section II, or may be derived from shorter sequence regions within cloned cDNA fragments. Shorter fragments can be prepared by enzymatic digestion of full-length fragments under conditions which yield desired-sized fragments, as will be described in Section IV. Alternatively, the fragments can be produced by oligonucleotide synthetic methods, using sequences derived from the cDNA fragments. Methods or services for producing selected-sequence oligonucleotide fragments are available.

To confirm that a given ET-NANB fragment is in fact derived from the ET-NANB viral agent, the fragment can be shown to hybridize selectively with cDNA from infected sources. By way of illustration, to confirm that the 1.33 kb fragment in the pTZ-KF1(E1.1) plasmid is ET-NANB in origin, the fragment was excised from the pTZ-KF1(ET1.1) plasmid, purified, and radiolabeled by random labeling. The radiolabeled fragment was

hybridized with fractionated cDNAs from infected and non-infected sources to confirm that the probe reacts only with infected-source cDNAs. This method is illustrated in Example 4, where the above radiolabeled 1.33 kb fragment from pTZ-KF1(ET1.1) plasmid was examined for binding to cDNAs prepared from infected and non-infected sources. The infected sources are (1) bile from a cynomolgus monkey infected with a strain of virus derived from stool samples from human patients from Burma with known ET-NANB infections and (2) a viral agent derived from the stool sample of a human ET-NANB patient from Mexico. The cDNAs in each fragment mixture were first amplified by a linker/primer amplification method described in Example 4. Fragment separation was on agarose gel, followed by Southern blotting and then hybridization to bind the radiolabeled 1.33 kb fragment to the fractionated cDNAs. The lane containing cDNAs from the infected sources showed a smeared band of bound probe, as expected (cDNAs amplified by the linker/primer amplification method would be expected to have a broad range of sizes). No probe binding to the amplified cDNAs from the non-infected sources was observed. The results indicate that the 1.33 kb probe is specific for cDNA fragments associated with ET-NANB infection. This same type of study, using ET 1.1 as the probe, has demonstrated hybridization to ET-NANB samples collected from Tashkent, Somalia, Borneo and Pakistan. Secondly, the fact that the probe is specific for ET-NANB related sequences derived from different continents (Asia, Africa and North America) indicates the cloned ET-NANB Africa sequence is derived from a common ET-NANB virus or virus class responsible for ET-NANB hepatitis infection worldwide.

In a related confirmatory study, probe binding to fractionated genomic fragments prepared from human or cynomolgus monkey genomic DNA was examined. No

probe binding was observed to either genomic fraction, demonstrating that the ET-NANB fragment is not an endogenous human or cynomologus fragment.

Another confirmation of ET-NANB specific sequences in the fragments is the ability to express ET-NANB proteins from coding regions in the fragments. Section IV below discusses methods of protein expression using the fragments.

One important use of the ET-NANB-specific fragments is for identifying ET-NANB-derived cDNAs which contain additional sequence information. The newly identified cDNAs, in turn, yield new fragment probes, allowing further iterations until the entire viral genome is identified and sequenced. Procedures for identifying additional ET-NANB library clones and generating new probes therefrom generally follow the cloning and selection procedures described in Section II.

The fragments (and oligonucleotides prepared based on the sequences given above) are also useful as primers for a polymerase chain reaction method of detecting ET-NANB viral genomic material in a patient sample. This diagnostic method will be described in Section V below.

IV. ET-NANB Proteins

As indicated above, ET-NANB proteins can be prepared by expressing open reading-frame coding regions in ET-NANB fragments. In one preferred approach, the ET-NANB fragments used for protein expression are derived from cloned cDNAs which have been treated to produce desired-size fragments, and preferably random fragments with sizes predominantly between about 100 to about 300 base pairs. Example 5 describes the preparation of such fragments by DNAs digestion. Because it is desired to obtain peptide

antigens of between about 30 to about 100 amino acids, the digest fragments are preferably size fractionated, for example by gel electrophoresis, to select those in the approximately 100-300 basepair size range.

A. Expression Vector

The ET-NANB fragments are inserted into a suitable expression vector. One exemplary expression vector is lambda gt11, which contains a unique EcoRI insertion site 53 base pairs upstream of the translation termination codon of the beta-galactosidase gene. Thus, the inserted sequence will be expressed as a beta-galactosidase fusion protein which contains the N-terminal portion of the beta-galactosidase gene, the heterologous peptide, and optionally the C-terminal region of the beta-galactosidase peptide (the C-terminal portion being expressed when the heterologous peptide coding sequence does not contain a translation termination codon). This vector also produces a temperature-sensitive repressor (c1857) which causes viral lysogeny at permissive temperatures, e.g., 32°C, and leads to viral lysis at elevated temperatures, e.g., 42°C. Advantages of this vector include: (1) highly efficient recombinant generation, (2) ability to select lysogenized host cells on the basis of host-cell growth at permissive, but not non-permissive, temperatures, and (3) high levels of recombinant fusion protein production. Further, since phage containing a heterologous insert produces an inactive beta-galactosidase enzyme, phage with inserts can be readily identified by a beta-galactosidase colored-substrate reaction.

For insertion into the expression vector, the viral digest fragments may be modified, if needed, to contain selected restriction-site linkers, such as EcoRI linkers, according to conventional procedures. Example 1 illustrates methods for cloning the digest

fragments into lambda gt11, which includes the steps of blunt-ending the fragments, ligating with EcoRI linkers, and introducing the fragments into EcoRI-cut lambda gt11. The resulting viral genomic library may be checked to confirm that a relatively large (representative) library has been produced. This can be done, in the case of the lambda gt11 vector, by infecting a suitable bacterial host, plating the bacteria, and examining the plaques for loss of beta-galactosidase activity. Using the procedures described in Example 1, about 50% of the plaques showed loss of enzyme activity.

B. Peptide Antigen Expression

The viral genomic library formed above is screened for production of peptide antigen (expressed as a fusion protein) which is immunoreactive with antiserum from ET-NANB seropositive individuals. In a preferred screening method, host cells infected with phage library vectors are plated, as above, and the plate is blotted with a nitrocellulose filter to transfer recombinant protein antigens produced by the cells onto the filter. The filter is then reacted with the ET-NANB antiserum, washed to remove unbound antibody, and reacted with reporter-labeled, anti-human antibody, which becomes bound to the filter, in sandwich fashion, through the anti-ET-NANB antibody.

Typically phage plaques which are identified by virtue of their production of recombinant antigen of interest are re-examined at a relatively low density for production of antibody-reactive fusion protein. Several recombinant phage clones which produced immunoreactive recombinant antigen were identified in the procedure.

The selected expression vectors may be used for scale-up production, for purposes of recombinant protein purification. Scale-up production is carried

out using one of a variety of reported methods for (a) lysogenizing a suitable host, such as E. coli, with a selected lambda gt11 recombinant (b) culturing the transduced cells under conditions that yield high levels of the heterologous peptide, and (c) purifying the recombinant antigen from the lysed cells.

In one preferred method involving the above lambda gt11 cloning vector, a high-producer E. coli host, BNN103, is infected with the selected library phage and replica plated on two plates. One of the plates is grown at 32°C, at which viral lysogeny can occur, and the other at 42°C, at which the infecting phage is in a lytic stage and therefore prevents cell growth. Cells which grow at the lower but not the higher temperature are therefore assumed to be successfully lysogenized.

The lysogenized host cells are then grown under liquid culture conditions which favor high production of the fused protein containing the viral insert, and lysed by rapid freezing to release the desired fusion protein.

C. Peptide Purification

The recombinant peptide can be purified by standard protein purification procedures which may include differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel electrophoresis and affinity chromatography. In the case of a fused protein, such as the beta-galactosidase fused protein prepared as above, the protein isolation techniques which are used can be adapted from those used in isolation of the native protein. Thus, for isolation of a beta-glactosidase fusion protein, the protein can be isolated readily by simple affinity chromatography, by passing the cell lysis material over a solid support having surface-bound anti-beta-galactosidase antibody.

D. Viral Proteins

The ET-NANB protein of the invention may also be derived directly from the ET-NANB viral agent. VLPs isolated from a stool sample from an infected individual, as above, are one suitable source of viral protein material. The VLPs isolated from the stool sample may be further purified by affinity chromatography prior to protein isolation (see below). The viral agent may also be raised in cell culture, which provides a convenient and potentially concentrated source of viral protein. Co-owned U.S. Patent Application Serial No. 846,757, filed April 1, 1986, describes an immortalized trioma liver cell which supports NANB infection in cell culture. The trioma cell line is prepared by fusing human liver cells with a mouse/human fusion partner selected for human chromosome stability. Cells containing the desired NANB viral agent can be identified by immunofluorescence methods, employing anti-ET-NANB human antibodies.

The viral agent is disrupted, prior to protein isolation, by conventional methods, which can include sonication, high- or low-salt conditions, or use of detergents.

Purification of ET-NANB viral protein can be carried out by affinity chromatography, using a purified anti-ET-NANB antibody attached according to standard methods to a suitable solid support. The antibody itself may be purified by affinity chromatography, where an immunoreactive recombinant ETNANB protein, such as described above, is attached to a solid support, for isolation of anti-ET-NANB antibodies from an immune serum source. The bound antibody is released from the support by standard methods.

Alternatively, the anti-ET-NANB antibody may be a monoclonal antibody (Mab) prepared by immunizing a mouse or other animal with recombinant ET-NANB protein, isolating lymphocytes from the animal and immortalizing the cells with a suitable fusion partner, and selecting successful fusion products which react with the recombinant protein immunogen. These in turn may be used in affinity purification procedures, described above, to obtain native ET-NANB antigen.

V. Utility

A. Diagnostic Methods

The particles and antigens of the invention, as well as the genetic material, can be used in diagnostic assays. Methods for detecting the presence of ET-NANB hepatitis comprise analyzing a biological sample such as a blood sample, stool sample or liver biopsy specimen for the presence of an analyte associated with ET-NANB hepatitis virus.

The analyte can be a nucleotide sequence which hybridizes with a probe comprising a sequence of at least about 16 consecutive nucleotides, usually 30 to 200 nucleotides, up to substantially the full sequence of the sequences shown above (cDNA sequences). The analyte can be RNA or cDNA. The analyte is typically a virus particle suspected of being ET-NANB or a particle for which this classification is being ruled out. The virus particle can be further characterized as having an RNA viral genome comprising a sequence at least about 80% homologous to a sequence of at least 12 consecutive nucleotides of the "forward" and "reverse" sequences given above, usually at least about 90% homologous to at least about 60 consecutive nucleotides within the sequences, and may comprise a sequence substantially homologous to the full-length sequences. In order to detect an analyte, where the

analyte hybridizes to a probe, the probe may contain a detectable label.

The analyte can also comprise an antibody which recognizes an antigen, such as a cell surface antigen, on a ET-NANB virus particle. The analyte can also be a ET-NANB viral antigen. Where the analyte is an antibody or an antigen, either a labelled antigen or antibody, respectively, can be used to bind to the analyte to form an immunological complex, which can then be detected by means of the label.

Typically, methods for detecting analytes such as surface antigens and/or whole particles are based on immunoassays. Immunoassays can be conducted either to determine the presence of antibodies in the host that have arisen from infection by ET-NANB hepatitis virus or by assays that directly determine the presence of virus particles or antigens. Such techniques are well known and need not be described here in detail. Examples include both heterogeneous and homogeneous immunoassay techniques. Both techniques are based on the formation of an immunological complex between the virus particle or its antigen and a corresponding specific antibody. Heterogeneous assays for viral antigens typically use a specific monoclonal or polyclonal antibody bound to a solid surface. Sandwich assays are becoming increasingly popular. Homogeneous assays, which are carried out in solution without the presence of a solid phase, can also be used, for example by determining the difference in enzyme activity brought on by binding of free antibody to an enzyme-antigen conjugate. A number of suitable assays are disclosed in U.S. Patent Nos. 3,817,837, 4,006,360, 3,996,345.

When assaying for the presence of antibodies induced by ET-NANB viruses, the viruses and antigens of the invention can be used as specific binding agents to detect either IgG or IgM antibodies. Since IgM antibodies are typically the first antibodies that appear

during the course of an infection, when IgG synthesis may not yet have been initiated, specifically distinguishing between IgM and IgG antibodies present in the blood stream of a host will enable a physician or other investigator to determine whether the infection is recent or chronic.

In one diagnostic configuration, test serum is reacted with a solid phase reagent having surface-bound ET-NANB protein antigen. After binding anti-ET-NANB antibody to the reagent and removing unbound serum components by washing, the reagent is reacted with reporter-labeled anti-human antibody to bind reporter to the reagent in proportion to the amount of bound anti-ET-NANB antibody on the solid support. The reagent is again washed to remove unbound labeled antibody, and the amount of reporter associated with the reagent is determined. Typically, the reporter is an enzyme which is detected by incubating the solid phase in the presence of a suitable fluorometric or colorimetric substrate.

The solid surface reagent in the above assay prepared by known techniques for attaching protein material to solid support material, such as polymeric beads, dip sticks, or filter material. These attachment methods generally include non-specific adsorption of the protein to the support or covalent attachment of the protein, typically through a free amine group, to a chemically reactive group on the solid support, such as an activate carboxyl, hydroxyl, or aldehyde group.

In a second diagnostic configuration, known as a homogeneous assay, antibody binding to a solid support produces some change in the reaction medium which can be directly detected in the medium. Known general types of homogeneous assays proposed heretofore include (a) spin-labeled reporters, where antibody binding to the antigen is detected by a change in

reported mobility (broadening of the spin splitting peaks), (b) fluorescent reporters, where binding is detected by a change in fluorescence efficiency, (c) enzyme reporters, where antibody binding effects enzyme/substrate interactions, and (d) liposome-bound reporters, where binding leads to liposome lysis and release of encapsulated reporter. The adaptation of these methods to the protein antigen of the present invention follows conventional methods for preparing homogeneous assay reagents.

In each of the assays described above, the assay method involves reacting the serum from a test individual with the protein antigen and examining the antigen for the presence of bound antibody. The examining may involve attaching a labeled anti-human antibody to the antibody being examined, either IgM (acute phase) or IgG (convalescent phase), and measuring the amount of reporter bound to the solid support, as in the first method, or may involve observing the effect of antibody binding on a homogeneous assay reagent, as in the second method.

Also forming part of the invention is an assay system or kit for carrying out the assay method just described. The kit generally includes a support with surface-bound recombinant protein antigen which is (a) immunoreactive with antibodies present in individuals infected with enterically transmitted nonA/nonB viral agent and (b) derived from a viral hepatitis agent whose genome contains a region which is homologous to the 1.33 kb DNA EcoRI insert present in plasmid pTZ-KF1(ET1.1) carried in E. Coli strain BB4, and having ATCC deposit no. 67717. A reporter-labeled anti-human antibody in the kit is used for detecting surface-bound anti-ET-NANB antibody.

B. Viral Genome Diagnostic Applications

The genetic material of the invention can itself be used in numerous assays as probes for genetic material present in naturally occurring infections. One method for amplification of target nucleic acids, for later analysis by hybridization assays, is known as the polymerase chain reaction or PCR technique. The PCR technique can be applied to detecting virus particles of the invention in suspected pathological samples using oligonucleotide primers spaced apart from each other and based on the genetic sequence set forth above. The primers are complementary to opposite strands of a double stranded DNA molecule and are typically separated by from about 50 to 450 nt or more (usually not more than 2000 nt). This method entails preparing the specific oligonucleotide primers and then repeated cycles of target DNA denaturation, primer binding, and extension with a DNA polymerase to obtain DNA fragments of the expected length based on the primer spacing. Extension products generated from one primer serve as additional target sequences for the other primer. The degree of amplification of a target sequence is controlled by the number of cycles that are performed and is theoretically calculated by the simple formula 2^n where n is the number of cycles. Given that the average efficiency per cycle ranges from about 65% to 85%, 25 cycles produce from 0.3 to 4.8 million copies of the target sequence. The PCR method is described in a number of publications, including Saiki <u>et al.</u>, <u>Science</u> (1985) <u>230</u>:1350-1354; Saiki <u>et al.</u>, <u>Nature</u> (1986) <u>324</u>:163-166; and Scharf <u>et al.</u>, <u>Science</u> (1986) <u>233</u>:1076-1078. Also see U.S. Patent Nos. 4,683,194; 4,683,195; and 4,683,202.

The invention includes a specific diagnostic method for determination of ET-NANB viral agent, based on selective amplification of ET-NANB fragments. This method employs a pair of single-strand primers derived

from non-homologous regions of opposite strands of a DNA duplex fragment, which in turn is derived from an enterically transmitted viral hepatitis agent whose genome contains a region which is homologous to the 1.33 kb DNA EcoRI insert present in plasmid pTZ-KF1(ET1.1) carried in E. coli strain BB4, and having ATCC deposit no. 67717. These "primer fragments," which form one aspect of the invention, are prepared from ET-NANB fragments such as described in Section III above. The method follows the process for amplifying selected nucleic acid sequences as disclosed in U.S. Patent No. 4,683,202, as discussed above.

C. Peptide Vaccine

Any of the antigens of the invention can be used in preparation of a vaccine. A preferred starting material for preparation of a vaccine is the particle antigen isolated from bile. The antigens are preferably initially recovered as intact particles as described above. However, it is also possible to prepare a suitable vaccine from particles isolated from other sources or non-particle recombinant antigens. When non-particle antigens are used (typically soluble antigens), proteins derived from the viral envelope or viral capsid are preferred for use in preparing vaccines. These proteins can be purified by affinity chromatography, also described above.

If the purified protein is not immunogenic *per se*, it can be bound to a carrier to make the protein immunogenic. Carriers include bovine serum albumin, keyhole limpet hemocyanin and the like. It is desirable, but not necessary, to purify antigens to be substantially free of human protein. However, it is more important that the antigens be free of proteins, viruses, and other substances not of human origin that may have been introduced by way of, or contamination of, the nutrient medium, cell lines, tissues, or patho-

logical fluids from which the virus is cultured or obtained.

Vaccination can be conducted in conventional fashion. For example, the antigen, whether a viral particle or a protein, can be used in a suitable diluent such as water, saline, buffered salines, complete or incomplete adjuvants, and the like. The immunogen is administered using standard techniques for antibody induction, such as by subcutaneous administration of physiologically compatible, sterile solutions containing inactivated or attenuated virus particles or antigens. An immune response producing amount of virus particles is typically administered per vaccinizing injection, typically in a volume of one milliliter or less.

A specific example of a vaccine composition includes, in a pharmacologically acceptable adjuvant, a recombinant protein or protein mixture derived from an enterically transmitted nonA/nonB viral hepatitis agent whose genome contains a region which is homologous to the 1.33 kb DNA EcoRI insert present in plasmid pTZ-KF1(ET1.1) carried in E. coli strain BB4, and having ATCC deposit no. 67717. The vaccine is administered at periodic intervals until a significant titer of anti-ET-NANB antibody is detected in the serum. The vaccine is intended to protect against ET-NANB infection.

D. Prophylactic and Therapeutic Antibodies and Antisera

In addition to use as a vaccine, the compositions can be used to prepare antibodies to ET-NANB virus particles. The antibodies can be used directly as antiviral agents. To prepare antibodies, a host animal is immunized using the virus particles or, as appropriate, non-particle antigens native to the virus particle are bound to a carrier as described above for vaccines. The host serum or plasma is collected

following an appropriate time interval to provide a composition comprising antibodies reactive with the virus particle. The gamma globulin fraction or the IgG antibodies can be obtained, for example, by use of saturated ammonium sulfate or DEAE Sephadex, or other techniques known to those skilled in the art. The antibodies are substantially free of many of the adverse side effects which may be associated with other anti-viral agents such as drugs.

The antibody compositions can be made even more compatible with the host system by minimizing potential adverse immune system responses. This is accomplished by removing all or a portion of the Fc portion of a foreign species antibody or using an antibody of the same species as the host animal, for example, the use of antibodies from human/human hybridomas.

The antibodies can also be used as a means of enhancing the immune response since antibody-virus complexes are recognized by macrophages. The antibodies can be administered in amounts similar to those used for other therapeutic administrations of antibody. For example, pooled gamma globulin is administered at 0.02-0.1 ml/lb body weight during the early incubation of other viral diseases such as rabies, measles and hepatitis B to interfere with viral entry into cells. Thus, antibodies reactive with the ET-NANB virus particle can be passively administered alone or in conjunction with another anti-viral agent to a host infected with an ET-NANB virus to enhance the immune response and/or the effectiveness of an antiviral drug.

Alternatively, anti-ET-NANB-virus antibodies can be induced by administering anti-idiotype antibodies as immunogens. Conveniently, a purified antiET-NANB-virus antibody preparation prepared as descibed above is used to induce anti-idiotype antibody in a host animal. The composition is administered to the host animal in a suitable diluent. Following

administration, usually repeated administration, the host produces anti-idiotype antibody. To eliminate an immunogenic response to the Fc region, antibodies produced by the same species as the host animal can be used or the Fc region of the administered antibodies can be removed. Following induction of anti-idiotype antibody in the host animal, serum or plasma is removed to provide an antibody composition. The composition can be purified as described above for anti-ET-NANB-virus antibodies, or by affinity chromatography using anti-ET-NANB-virus antibodies bound to the affinity matrix. The anti-idiotype antibodies produced are similar in conformation to the authentic ET-NANB antigen and may be used to prepare an ET-NANB vaccine rather than using a ET-NANB particle antigen.

When used as a means of inducing anti-ET-NANB-virus antibodies in a patient, the manner of injecting the antibody is the same as for vaccination purposes, namely intramuscularly, intraperitoneally, subcutaneously or the like in an effective concentration in a physiologically suitable diluent with or without adjuvant. One or more booster injections may be desirable. The anti-idiotype method of induction of anti-ET-NANB-virus antibodies can alleviate problems which may be caused by passive administration of anti-ET-NANB-virus antibodies, such as an adverse immune response, and those associated with administration of purified blood components, such as infection with as yet undiscovered viruses.

The ET-NANB derived proteins of the invention are also intended for use in producing antiserum designed for pre- or post-exposure prophylaxis. Here an ET-NANB protein, or mixture of proteins is formulated with a suitable adjuvant and administered by injection to human volunteers, according to known methods for producing human antisera. Antibody response to the injected proteins is monitored, during

a several- week period following immunization, by periodic serum sampling to detect the presence an anti-ET-NANB serum antibodies, as described in Section IIA above.

The antiserum from immunized individuals may be administered as a pre-exposure prophylactic measure for individuals who are at risk of contracting infection. The anitserum is also useful in treating an individual post-exposure, analogous to the use of high titer antiserum against hepatitis B virus for post-exposure prophylaxis.

E. Monoclonal Antibodies

For both *in vivo* use of antibodies to ET-NANB virus particles and proteins and anti-idiotype antibodies and diagnostic use, it may be preferable to use monoclonal antibodies. Monoclonal anti-virus particle antibodies or anti-idiotype antibodies can be produced as follows. The spleen or lymphocytes from an immunized animal are removed and immortalized or used to prepare hybridomas by methods known to those skilled in the art. To produce a human-human hybridoma, a human lymphocyte donor is selected. A donor known to be infected with a ET-NANB virus (where infection has been shown for example by the presence of anti-virus antibodies in the blood or by virus culture) may serve as a suitable lymphocyte donor. Lymphocytes can be isolated from a peripheral blood sample or spleen cells may be used if the donor is subject to splenectomy. Epstein-Barr virus (EBV) can be used to immortalize human lymphocytes or a human fusion partner can be used to produce human-human hybridomas. Primary *in vitro* immunization with peptides can also be used in the generation of human monoclonal antibodies.

Antibodies secreted by the immortalized cells are screened to determine the clones that secrete antibodies of the desired specificity. For monoclonal

anti-virus particle antibodies, the antibodies must bind to ET-NANB virus particles. For monoclonal anti-idiotype antibodies, the antibodies must bind to anti-virus particle antibodies. Cells producing antibodies of the desired specificity are selected.

The following examples illustrate various aspects of the invention, but are in no way intended to limit the scope thereof.

Material

The materials used in the following Examples were as follows:

Enzymes: DNAse I and alkaline phosphatase were obtained from Boehringer Mannheim Biochemicals (BMB, Indianapolis, IN); EcoRI, EcoRI methylase, DNA ligase, and DNA Polymerase I, from New England Biolabs (NEB, Beverly MA); and RNase A was obtained from Sigma (St. Louis, MO).

Other reagents: EcoRI linkers were obtained from NEB; and nitro blue tetrazolium (NBT), 5-bromo-4-chloro-3-indolyl phosphate (BCIP) 5-bromo-4-chloro-3-indolyl-B-D-galactopyranoside (X-gal) and isopropyl B-D-thiogalactopyranoside (IPTG) were obtained from Sigma.

cDNA synthesis kit and random priming labeling kits are available from Boehringer-Mannheim Biochemical (BMB, Indianapolis, IN).

Eample 1
Preparing cDNA Library

A. Source of ET-NANB virus

Two cynomolgus monkeys (cynos) were intravenously injected with a 10% suspension of a stool pool obtained from a second-passage cyno (cyno #37) infected with a strain of ET-NANB virus isolated from Burma cases whose stools were positive for ET-NANB, as

evidenced by binding of 27-34 nm virus-like particles (VLPs) in the stool to immune serum from a known ET-NANB patient. The animals developed elevated levels of alanine aminotransferase (ALT) between 24-36 days after innoculation, and one excreted 27-34 nm VLPs in its bile in the pre-acute phase of infection.

The bile duct of each infected animal was cannulated and about 1-3 cc of bile was collected daily. RNA was extracted from one bile specimen (cyno #121) by hot phenol extraction, using a standard RNA isolation procedure. Double-strand cDNA was formed from the isolated RNA by a random primer for first-strand generation, using a cDNA synthesis kit obtained from Boehringer-Mannheim (Indianapolis, IN).

B. Cloning the Duplex Fragments

The duplex cDNA fragments were blunt-ended with T4 DNA polymerase under standard conditions (Maniatis, p. 118), then extracted with phenol/chloroform and precipitated with ethanol. The blunt-ended material was ligated with <u>Eco</u>RI linkers under standard conditions (Maniatis, pp. 396-397) and digested with <u>Eco</u>RI to remove redundant linker ends. Non-ligated linkers were removed by sequential isopropanol precipitation.

Lambda gt10 phage vector (Huynh) was obtained from Promega Biotec (Madison, WI). This cloning vector has a unique <u>Eco</u>RI cloning site in the phage cI repressor gene. The cDNA fragments from above were introduced into the <u>Eco</u>RI site by mixing 0.5 - 1.0 ug <u>Eco</u>RI-cleaved gt10, 0.5-3 µl of the above duplex fragments, 0.5 µl 10X ligation buffer, 0.5 µl ligase (200 units), and distilled water to 5 µl. The mixture was incubated overnight at 14°C, followed by <u>in vitro</u> packaging, according to standard methods (Maniatis, pp. 256-268).

The packaged phage were used to infect an E. coli hfl strain, such as strain HG415. Alternatively, E. coli, strain C600 hfl, avialable from Promega Biotec, Madison, WI, could be used. The percentage of recombinant plaques obtained with insertion of the EcoRI-ended fragments was less than 5% by analysis of 20 random plaques.

The resultant cDNA library was plated and phage were eluted from the selection plates by addition of elution buffer. After DNA extraction from the phage, the DNA was digested with EcoRI to release the heterogeneous insert population, and the DNA fragments were fractionated on agarose to remove phage fragments. The 500-4,000 basepair inserts were isolated and recloned into lambda gt10 as above, and the packaged phage was used to infect E. coli strain HG415. The percentage of successful recombinants was greater than 95%. The phage library was plated on E. coli strain HG415, at about 5,000 plaques/plate, on a total of 8 plates.

Example 2

Selecting ET-NANB Cloned Fragments

A. cDNA Probes

Duplex cDNA fragments from noninfected and ET-NANB-infected cynomolgus monkeys were prepared as in Example 1. The cDNA fragments were radiolabeled by random priming, using a random-priming labeling kit obtained from Boehringer-Mannheim (Indianapolis, IN).

B. Clone Selection

The plated cDNA library from Example I was transferred to each of two nitrocellulose filters, and the phage DNA was fixed on the filters by baking, according to standard methods (Maniatis, pp. 320-323). The duplicate filters were hybridized with either infected-source or control cDNA probes from above. Autoradiographs of the filters were examined to identify library clones which hybridized with radiolabeled cDNA probes from infected source only, i.e., did not hybridize with cDNA probes from the non-infected source. Sixteen such clones, out of a total of about 40,000 clones examined, were identified by this subtraction selection method.

Each of the sixteen clones was picked and replated at low concentration on an agar plate. The clones on each plate were transferred to two nitrocellulose as duplicate lifts, and examined for hybridization to radiolabeled cDNA probes from infected and noninfected sources, as above. Clones were selected which showed selective binding for infected-source probes (i.e., binding with infected-source probes and substantially no binding with non-infected-source probes). One of the clones which bound selectively to probe from infected source was isolated for further study. The selected vector was identified as lambda gt10-1.1, indicated in Figure 1.

Example 3

ET-NANB Sequence

Clone lambda gt10-1.1 from Example 2 was digested with EcoRI to release the heterologous insert, which was separated from the vector fragments by gel electrophoresis. The electrophoretic mobility of the fragment was consistent with a 1.33 kb fragment. This fragment, which contained EcoRI ends, was inserted into the EcoRI site of a pTZ-KF1 vector, whose construction and properties are described in co-owned U.S. patent application for "Cloning Vector System and Method for Rare Clone Identification", Serial No. 125, 650, filed November 25, 1987. Briefly, and as illustrated in Figure 1, this plasmid contains a unique EcoRI site adjacent a T7 polymerase promoter site, and plasmid and phage origins of replication. The sequence immediately adjacent each side of the EcoRI site is known. *E. coli* BB4 bacteria, obtained from Stratagene (La Jolla, CA, were transformed with the plasmid.

Radiolabeled ET-NANB probe was prepared by excising the 1.33 kb insert from the lambda gt10-1.1 phage in Example 2, separating the fragment by gel electrophoresis, and randomly labeling as above. Bacteria transfected with the above pTZ-KF1 and containing the desired ET-NANB insert were selected by replica lift and hybridization with the radiolabeled ET-NANB probe, according to methods outlined in Example 2.

One bacterial colony containing a successsful recombinant was used for sequencing a portion of the 1.33 kb insert. This isolate, designated pTZ-KF1(ET1.1), has been deposited with the American Type Culture Collection, and is identified by ATCC deposit no. 67717. Using a standard dideoxy sequencing procedure, and primers for the sequences flanking the EcoRI site, about 200-250 basepairs of sequence from

the 5'-end region and 3'-end region of the insert were obtained. The sequences are given above in Section II. Later sequencing by the same techniques gave the full sequence in both directions, also given above.

Example 4

Detecting ET-NANB Sequences

cDNA fragment mixtures from the bile of noninfected and ET-NANB-infected cynomolgus monkeys were prepared as above. The cDNA fragments obtained from human stool samples were prepared as follows. Thirty ml of a 10% stool suspension obtained from an individual from Mexico diagnosed as infected with ET-NANB as a result of an ET-NANB outbreak, and a similar volume of stool from a healthy, non-infected individual, were layered over a 30% sucrose density gradient cushion, and centrifuged at 25,000 xg for 6 hr in an SW27 rotor, at 15°C. The pelleted material from the infected-source stool contained 27-34 nm VLP particles characteristic of ET-NANB infection in the infected-stool sample. RNA was isolated from the sucrose-gradient pellets in both the infected and non-infected samples, and the isolated RNA was used to produce cDNA fragments as described in Example 1.

The cDNA fragment mixtures from infected and non-infected bile source, and from infected and non-infected human-stool source were each amplified by a novel linker/primer replication method described in co-owned patent application serial number 07/208,512 for "DNA Amplification and Subtraction Technique," filed June 17, 1988. Briefly, the fragments in each sample were blunt-ended with DNA Pol I then extracted with phenol/chloroform and precipitated with ethanol. The blunt-ended material was ligated with linkers having the following sequence:

5'-GGAATTCGCGGCCGCTCG-3'
3'-TTCCTTAAGCGCCGGCGAGC-5'

The duplex fragements were digested with NruI to remove linker dimers, mixed with a primer having the sequence 5'-GGAATTCGCGGCCGCTCG-3', and then heat denatured and cooled to room temperature to form single-strand DNA/primer complexes. The complexes were replicated to form duplex fragments by addition of Thermus aquaticus (Taq) polymerase and all four deoxynucleotides. The replication procedures, involving successive strand denaturation, formation of strand/primer complexes, and replication, was repeated 25 times.

The amplified cDNA sequences were fractionated by agarose gel electrophoresis, using a 2% agarose matrix. After transfer of the DNA fragments from the agarose gels to nitrocellulose paper, the filters were hybridized to a random-labeled ^{32}P probe prepared by (i) treating the pTZ-KF1(ET1.1) plasmid from above with EcoRI, (ii) isolating the released 1.33 kb ET-NANB fragment, and (iii) randomly labeling the isolated fragment. The probe hybridization was performed by conventional Southern blotting methods (Maniatis, pp. 382-389). Figure 2 shows the hybridization pattern obtained with cDNAs from infected (I) and non-infected (N) bile sources (2A) and from infected (I) and non-infected (N) human stool sources (2B). As seen, the ET-NANB probe hybridized with fragments obtained from both of the infected sources, but was non-homologous to sequences obtained from either of the non-infected sources, thus confirming the specificity of derived sequence.

Southern blots of the radiolabeled 1.33 kb fragment with genomic DNA fragments from both human and cynomologus-monkey DNA were also prepared. No probe hybridization to either of the genomic fragment

mixtures was observed, confirming that the ET-NANB sequence is exogenous to either human or cynomolgus genome.

Example 5
Expressing ET-NANB Proteins

A. Preparing ET-NANB Coding Sequences

The pTZ-KF1(ET1.1) plasmid from Example 2 was digested with EcoRI to release the 1.33 kb ET-NANB insert which was purified from the linearized plasmid by gel electrophoresis. The purified fragment was suspended in a standard digest buffer (0.5M Tris HCl, pH 7.5; 1 mg/ml BSA; 10mM $MnCl_2$) to a concentration of about 1 mg/ml and digested with DNAse I at room temperature for about 5 minutes. These reaction conditions were determined from a prior calibration study, in which the incubation time required to produce predominantly 100-300 basepair fragments was determined. The material was extracted with phenol/chloroform before ethanol precipitation.

The fragments in the digest mixture were blunt-ended and ligated with EcoRI linkers as in Example 1. The resultant fragments were analyzed by electrophoresis (5-10V/cm) on 1.2% agarose gel, using PhiX174/HaeIII and lambda/HindIII size markers. The 100-300 bp fraction was eluted onto NA45 strips (Schleicher and Schuell), which were then placed into 1.5 ml microtubes with eluting solution (1 M NaCl, 50 mM arginine, pH 9.0), and incubated at 67°C for 30-60 minutes. The eluted DNA was phenol/chloroform extracted and then precipitated with two volumes of ethanol. The pellet was resuspended in 20 µl TE (0.01 M Tris HCl, pH 7.5, 0.001 M EDTA).

B. Cloning in an Expression Vector

Lambda gt11 phage vector (Huynh) was obtained from Promega Biotec (Madison, WI). This cloning vector has a unique EcoRI cloning site 53 base pairs upstream from the beta-galactosidase translation termination codon. The genomic fragments from above were introduced into the EcoRI site by mixing 0.5-1.0 µg EcoRI-cleaved gt11, 0.3-3 µl of the above sized fragments, 0.5 µl 10X ligation buffer (above), 0.5 µl ligase (200 units), and distilled water to 5 µl. The mixture was incubated overnight at 14°C, followed by *in vitro* packaging, according to standard methods (Maniatis, pp. 256-268).

The packaged phage were used to infect *E. coli* strain KM392, obtained from Dr. Kevin Moore, DNAX (Palo Alto, CA). Alternatively, *E. Coli* strain Y1090, available from the American Type Culture Collection (ATCC #37197), could be used. The infected bacteria were plated and the resultant colonies were checked for loss of beta-galactosidase activity-(clear plaques) in the presence of X-gal using a standard X-gal substrate plaque assay method (Maniatis). About 50% of the phage plaques showed loss of beta-galactosidase enzyme activity (recombinants).

C. Screening for ET-NANB Recombinant Proteins

ET-NANB convalescent antiserum was obtained from patients infected during documented ET-NANB outbreaks in Mexico, Borneo, Pakistan, Somalia, and Burma. The sera were immunoreactive with VLPs in stool specimens from each of several other patients with ET-NANB hepatitis.

A lawn of *E. coli* KM392 cells infected with about 10^4 pfu of the phage stock from above was prepared on a 150 mm plate and incubated, inverted, for 5-8 hours at 37°C. The lawn was overlaid with a nitrocellulose sheet, causing transfer of expressed ET-

NANB recombinant protein from the plaques to the paper. The plate and filter were indexed for matching corresponding plate and filter positions.

The filter was washed twice in TBST buffer (10 mM Tris, pH 8.0, 105 mM NaCl, 0.05% Tween 20), blocked with AIB (TBST buffer with 1% gelatin), washed again in TBST, and incubated overnight after addition of antiserum (diluted to 1:50 in AIB, 12-15 ml/plate). The sheet was washed twice in TBST and then contacted with enzyme-labeled anti-human antibody to attach the labeled antibody at filter sites containing antigen recognized by the antiserum. After a final washing, the filter was developed in a substrate medium containing 33 µl NBT (50 mg/ml stock solution maintained at 5°C) mixed with 16 µl BCIP (50 mg/ml stock solution maintained at 5°C) in 5 ml of alkaline phosphatase buffer (100 mM Tris, 9.5, 100 mM NaCl, 5 mM $MgCl_2$). Purple color appeared at points of antigen production, as recognized by the antiserum.

D. Screening Plating

The areas of antigen production determined in the previous step were replated at about 100-200 pfu on an 82 mm plate. The above steps, beginning with a 5-8 hour incubation, through NBT-BCIP development, were repeated in order to plaque purify phage secreting an antigen capable of reacting with the ET-NANB antibody. The identified plaques were picked and eluted in phage buffer (Maniatis, p. 443).

E. Epitope Identification

A series of subclones derived from the original pTZ-KF1 (ET1.1) plasmid from Example 2 were isolated using the same techniques described above. Each of these five subclones were immunoreactive with a pool of anti-ET antisera noted in C. The subclones contained short sequences from the "reverse" sequence set forth

previously. The beginning and ending points of the sequences in the subclones (relative to the full "reverse" sequence), are identified in the table below.

TABLE 1

Subclone	Position in "Reverse" Sequence	
	5'-end	3'-end
Y1	522	643
Y2	594	667
Y3	508	665
Y4	558	752
Y5	545	665

Since all of the gene sequences identified in the table must contain the coding sequence for the epitope, it is apparent that the coding sequence for the epitope falls in the region between nucleotide 594 (5'-end) and 643 (3'-end). Genetic sequences equivalent to and complementary to this relatively short sequence are therefore particularly preferred aspects of the present invention, as are peptides produced using this coding region.

A second series of clones identifying an altogether different epitope was isolated with only Mexican serum.

TABLE 2

Subclone	Position in "Forward" Sequence	
	5'end	3' end
ET 2-2	2	193
ET 8-3	2	135
ET 9-1	2	109
ET 13-1	2	101

The coding system for this epitope falls between nucleotide 2 (5'-end) and 101 (3'-end). Genetic sequences related to this short sequence are therefore also preferred, as are peptides produced using this coding region.

While the invention has been described with reference to particular embodiments, methods, construction and use, it will be apparent to those skilled in the art that various changes and modifications can be made without departing from the invention.

WE CLAIM:

1. A protein derived from an enterically transmitted non-A/non-B viral hepatitis agent whose genome contains a region which is homologous to the 1.33 kb DNA EcoRI insert present in plasmid pTZ-KF1(ET1.1) carried in *E. coli* strain BB4 and having ATCC deposit no. 67717.

2. The protein of claim 1, which is encoded by a coding region within said 1.33 kb EcoRI insert.

3. A recombinant protein derived from an enterically transmitted nonA/nonB viral hepatitis agent whose genome contains a region which is homologous to a duplex DNA having a first sequence:

```
1         11        21        31        41        51
*         *         *         *         *         *
CGAATTCCGCCAACTGATGGAAGGCACTAATCTGGCAAGACCTGTCCCTGTTGCAGCTGT

61        71        81        91        101       111
*         *         *         *         *         *
TCTACCACCCTGCCCCGAGCTCGAACAGGGCCTTCTCTACCTGCCCCAGGAGCTCACCAC

121       131       141       151       161       171
*         *         *         *         *         *
CTGTGATAGTGTCGTAACATTTGAATTAACAGACATTGTGCACTGCCGCATGGCCGCCCC
```

```
181       191       201       211       221       231
*         *         *         *         *         *
GAGCCAGCGCAAGGCCGTGCTGTCCACACTCGTGGGCCGCTACGGCGTCGCACAAAGCTC

241       251       261       271       281       291
*         *         *         *         *         *
TACAATGCTTCCCACTCTGATGTTCGCGACTCTCTCGCCCGTTTTATCCCGGCCATTGGC

301       311       321       331       341       351
*         *         *         *         *         *
CCCGTACAGGTTACAACTTGTGAATTGTACGAGCTAGTGGAGGCCATGGTCGAGAAGGGC

361       371       381       391       401       411
*         *         *         *         *         *
CAGGATGGCTCCGCCGTCCTTGAGCTTGATCTTTGCAACCGTGACGTGTCCAGGATCACC

421       431       441       451       461       471
*         *         *         *         *         *
TTCTTCCAGAAAGATTGTAACAAGTTCACCACAGGTGAGACCATTGCCCATGGTAAAGTG

481       491       501       511       521       531
*         *         *         *         *         *
GGCCAGGGCATCTCGGCCTGGAGCAAGACCTTCTGCGCCCTCTTTGGCCCTTGGTTCCGC

541       551       561       571       581       591
*         *         *         *         *         *
GCTATTGAGAAGGCTATTCTGGCCCTGCTCCCTCAGGGTGTGTTTTACGGTGATGCCTTT

601       611       621       631       641       651
*         *         *         *         *         *
GATGACACCGTCTTCTCGGCGGCTGTGGCCGCAGCAAAGGCATCCATGGTGTTTGAGAAT

661       671       681       691       701       711
*         *         *         *         *         *
GACTTTTCTGAGTTTGACTCCACCCAGAATAACTTTTCTCTGGGTCTAGAGTGTGCTATT
```

721 731 741 751 761 771
* * * * * *
ATGGAGGAGTGTGGGATGCCGCAGTGGCTCATCCGCCTGTATCACCTTATAAGGTCTGCG

781 791 801 811 821 831
* * * * * *
TGGATCTTGCAGGCCCCGAAGGAGTCTCTGCGAGGGTTTTGGAAGAAACACTCCGGTGAG

841 851 861 871 881 891
* * * * * *
CCCGGCACTCTTCTATGGAATACTGTCTGGAATATGGCCGTTATTACCCACTGTTATGAC

901 911 921 931 941 951
* * * * * *
TTCCGCGATTTTCAGGTGGCTGCCTTTAAAGGTGATGATTCGATAGTGCTTTGCAGTGAG

961 971 981 991 1001 1011
* * * * * *
TATCGTCAGAGTCCAGGAGCTGCTGTCCTGATCGCCGGCTGTGGCTTGAAGTTGAAGGTA

1021 1031 1041 1051 1061 1071
* * * * * *
GATTTCCGCCCGATCGGTTTGTATGCAGGTGTTGTGGTGGCCCCCGGCCTTGGCGCGCTC

1081 1091 1101 1111 1121 1131
* * * * * *
CCTGATGTTGTGCGCTTCGCCGGCCGGCTTACCGAGAAGAATTGGGGCCCTGGCCCTGAG

1141 1151 1161 1171 1181 1191
* * * * * *
CGGGCGGAGCAGCTCCGCCTCGCTGTTAGTGATTTCCTCCGCAAGCTCACGAATGTAGCT

1201 1211 1221 1231 1241 1251
* * * * * *
CAGATGTGTGTGGATGTTGTTTCCCGTGTTTATGGGGTTTCCCCTGGACTCGTTCATAAC

```
1261      1271      1281      1291      1301      1311
*         *         *         *         *         *
CTGATTGGCATGCTACAGGCTGTTGCTGATGGCAAGGCACATTTCACTGAGTCAGTAAAA

1321      1331      1341
*         *         *
CCAGTGCTCGACCGGAATTCGAGC,
```

a second sequence

```
1         11        21        31        41        51
*         *         *         *         *         *
GCTCGAATTCCGGTCGAGCACTGGTTTTACTGACTCAGTGAAATGTGCCTTGCCATCAGC

61        71        81        91        101       111
*         *         *         *         *         *
AACAGCCTGTAGCATGCCAATCAGGTTATGAACGAGTCCAGGGGAAACCCCATAAACACG

121       131       141       151       161       171
*         *         *         *         *         *
GGAAACAACATCCACACACATCTGAGCTACATTCGTGAGCTTGCGGAGGAAATCACTAAC

181       191       201       211       221       231
*         *         *         *         *         *
AGCGAGGCGGAGCTGCTCCGCCCGCTCAGGGCCAGGGCCCCAATTCTTCTCGGTAAGCCG

241       251       261       271       281       291
*         *         *         *         *         *
GCCGGCGAAGCGCACAACATCAGGGAGCGCGCCAAGGCCGGGGGCCACCACAACACCTGC

301       311       321       331       341       351
*         *         *         *         *         *
ATACAAACCGATCGGGCGGAAATCTACCTTCAACTTCAAGCCACAGCCGGCGATCAGGAC

361       371       381       391       401       411
*         *         *         *         *         *
AGCAGCTCCTGGACTCTGACGATACTCACTGCAAAGCACTATCGAATCATCACCTTTAAA
```

421 431 441 451 461 471
* * * * * *
GGCAGCCACCTGAAAATCGCGGAAGTCATAACAGTGGGTAATAACGGCCATATTCCAGAC

481 491 501 511 521 531
* * * * * *
AGTATTCCATAGAAGAGTGCCGGGCTCACCGGAGTGTTTCTTCCAAAACCCTCGCAGAGA

541 551 561 571 581 591
* * * * * *
CTCCTTCGGGGCCTGCAAGATCCACGCAGACCTTATAAGGTGATACAGGCGGATGAGCCA

601 611 621 631 641 651
* * * * * *
CTGCGGCATCCCACACTCCTCCATAATAGCACACTCTAGACCCAGAGAAAAGTTATTCTG

661 671 681 691 701 711
* * * * * *
GGTGGAGTCAAACTCAGAAAAGTCATTCTCAAACACCATGGATGCCTTTGCTGCGGCCAC

721 731 741 751 761 771
* * * * * *
AGCCGCCGAGAAGACGGTGTCATCAAAGGCATCACCGTAAAACACACCCTGAGGGAGCAG

781 791 801 811 821 831
* * * * * *
GGCCAGAATAGCCTTCTCAATAGCGCGGAACCAAGGGCCAAAGAGGGCGCAGAAGGTCTT

841 851 861 871 881 891
* * * * * *
GCTCCAGGCCGAGATGCCCTGGCCCACTTTACCATGGGCAATGGTCTCACCTGTGGTGAA

901 911 921 931 941 951
* * * * * *
CTTGTTACAATCTTTCTGGAAGAAGGTGATCCTGGACACGTCACGGTTGCAAAGATCAAG

```
961       971       981       991       1001      1011
*         *         *         *         *         *
CTCAAGGACGGCGGAGCCATCCTGGCCCTTCTCGACCATGGCCTCCACTAGCTCGTACAA

1021      1031      1041      1051      1061      1071
*         *         *         *         *         *
TTCACAAGTTGTAACCTGTACGGGGCCAATGGCCGGGATAAAACGGGCGAGAGAGTCGCG

1081      1091      1101      1111      1121      1131
*         *         *         *         *         *
AACATCAGAGTGGGAAGCATTGTAGAGCTTTGTGCGACGCCGTAGCGGCCCACGAGTGTG

1141      1151      1161      1171      1181      1191
*         *         *         *         *         *
GACAGCACGGCCTTGCGCTGGCTCGGGGCGGCCATGCGGCAGTGCACAATGTCTGTTAAT

1201      1211      1221      1231      1241      1251
*         *         *         *         *         *
TCAAATGTTACGACACTATCACAGGTGGTGAGCTCCTGGGGCAGGTAGAGAAGGCCCTGT

1261      1271      1281      1291      1301      1311
*         *         *         *         *         *
TCGAGCTCGGGGCAGGGTGGTAGAACAGCTGCAACAGGGACAGGTCTTGCCAGATTAGTG

1321      1331      1341
*         *         *
CCTTCCATCAGTTGGCGGAATTCG,
```

a third sequence

```
  1 CGGTTGTTCA GTACCAGTTT ACTGCAGGTG TGCCTGGATC CGGCAAGTCC
 51 CGCTCTATCA CCCAAGCCGA TGTGGACGTT GTCGTGGTCC CGACGCGTGA
101 GTTGCGTAAT GCCTGGCGCC GTCGCGGCTT TGCTGCTTTT ACCCCGCATA
151 CTGCCGCCAG AGTCACCCAG GGGCGCCGGG TTGTCATTGA TGAGGCTCCA
201 TCCCTCCCCC CTCACCTGCT GCTGCTCCAC ATGCAGCGGG CCGCCACCGT
251 CCACCTTCTT GGCGCCCCGA ACCAGATCCC AGCCATCGAC TTTGAGCACG
301 CTGGGCTCGT CCCCGCCATC AGGCCCGACT TAGCCCCACC TCCTGGTGGC
```

351 ATGTTACCCA TCGCTGCCCT GCGGATGTAT GCGAGCTCAT CCGTGGTGCA
401 TACCCCATGA TCCAGACCAC TAGCCGGGTT CTCCGTTCGT TGTTCTGGGG
451 TGAGCCTGCC GTCGGGCAGA AACTAGTGTT CACCCAGGCG GCCAAGGCCG
501 CCAACCCCGG CTCAGTGACG GTCCACGAGG CGCAGGGCGC TACCTACACG
551 GAGACCACTA TTATTGCCAC AGCAGATGCC CGGGGCCTTA TTCAGTCGTC
601 TCGGGCTCAT GCCATTGTTG CTCTGACGCG CCACACTGAG AAGTGCGTCA
651 TCATTGACGC ACCAGGCCTG CTTCGCGAGG TGGGCATCTC CGATGCAATC
701 GTTAATAACT TTTTCCTCGC TGGTGGCGAA ATTGGTCACC AGCGCCCATC
751 AGTTATTCCC CGTGGCAACC CTGACGCCAA TGTTGACACC CTGGCTGCCT
801 TCCCGCCGTC TTGCCAGATT AGTGCCTTCC ATCAGTTGGC TGAGGAGCTT
851 GGCCACAGAC CTGTCCCTGT TGCAGCTGTT CTACCACCCT GCCCCGAGCT
901 CGAACAGGGC CTTCTCTACC TGCCCCAGGA GCTCACCACC TGTGATAGTG
951 TCGTAACATT TGAATTAACA GACATTGTGC ACTGCCGCAT GGCCGCCCCG
1001 AGCCAGCGCA AGGCCGTGCT GTCCACACTC GTGGGCCGCT ACGGCGTCGC
1051 ACAAAGCTCT ACAATGCTTC CCACTCTGAT GTTCGCGACT CTCTCGCCCG
1101 TTTTATCCCG GCCATTGGCC CCGTACAGGT TACAACTTGT GAATTGTACG
1151 AGCTAGTGGA GGCCATGGTC GAGAAGGGCC AGGATGGCTC CGCCGTCCTT
1201 GAGCTTGATC TTTGCAACCG TGACGTGTCC AGGATCACCT TCTTCCAGAA
1251 AGATTGTAAC AAGTTCACCA CAGGTGAGAC CATTGCCCAT GGTAAAGTGG
1301 GCCAGGGCAT CTCGGCCTGG AGCAAGACCT TCTGCGCCCT CTTTGGCCCT
1351 TGGTTCCGCG CTATTGAGAA GGCTATTCTG GCCCTGCTCC CTCAGGGTGT
1401 GTTTTACGGT GATGCCTTTG ATGACACCGT CTTCTCGGCG GCTGTGGCCG
1451 CAGCAAAGGC ATCCATGGTG TTTGAGAATG ACTTTTCTGA GTTTGACTCC
1501 ACCCAGAATA ACTTTTCTCT GGGTCTAGAG TGTGCTATTA TGGAGGAGTG
1551 TGGGATGCCG CAGTGGCTCA TCCGCCTGTA TCACCTTATA AGGTCTGCGT
1601 GGATCTTGCA GGCCCCGAAG GAGTCTCTGC GAGGGTTTTG GAAGAAACAC
1651 TCCGGTGAGC CCGGCACTCT TCTATGGAAT ACTGTCTGGA ATATGGCCGT
1701 TATTACCCAC TGTTATGACT TCCGCGATTT TCAGGTGGCT GCCTTTAAAG
1751 GTGATGATTC GATAGTGCTT TGCAGTGAGT ATCGTCAGAG TCCAGGAGCT
1801 GCTGTCCTGA TCGCCGGCTG TGGCTTGAAG TTGAAGGTAG ATTTCCGCCC
1851 GATCGGTTTG TATGCAGGTG TTGTGGTGGC CCCCGGCCTT GGCGCGCTCC
1901 CTGATGTTGT GCGCTTCGCC GGCCGGCTTA CCGAGAAGAA TTGGGGCCCT
1951 GGCCCTGAGC GGGCGGAGCA GCTCCGCCTC GCTGTTAGTG ATTTCCTCCG
2001 CAAGCTCACG AATGTAGCTC AGATGTGTGT GGATGTTGTT TCCCGTGTTT
2051 ATGGGGTTTC CCCTGGACTC GTTCATAACC TGATTGGCAT GCTACAGGCT
2101 GTTGCTGATG GCAAGGCACA TTTCACTGAG TCAGTAAAAC CAGTGCTCGA
2151 CTTGACAAAT TCAATCTTGT GTCGGGTGGA ATGAATAACA TGTCTTTTGC

2201 TGCGCCCATG GGTTCGCGAC CATGCGCCCT CGGCCTATTT TGTTGCTGCT
2251 CCTCATGTTT TTGCCTATGC TGCCCGCGCC ACCGCCCGGT CAGCCGTCTG
2301 GCCGCCGTCG TGGGCGGCGC AGCGGCGGTT CCGGCGGTGG TTTCTGGGGT
2351 GACCGGGTTG ATTCTCAGCC CTTCGCAATC CCCTATATTC ATCCAACCAA
2401 CCCCTTCGCC CCCGATGTCA CCGCTGCGGC CGGGGCTGGA CCTCGTGTTC
2451 GCCAACCCGC CCGACCACTC GCGTCCGCTT GGCGTGACCA GGCCCAGCGC
2501 CCCGCCGTTG CCTCACGTCG TAGACCTACC ACAGCTGGGG CCGCGCCGCT
2551 AACCGCGGTC GCTCCGGCCC CG,

or a fourth sequence

1 GCGGCCGCTC GGTGGGGTTC CACCTCTCTT CCCATTCCGA ACAGAGAAGT
51 TAAGCAAGGA ATTAATTCGC GGCCGCTCGT GTTGCGTGAG GTGGGTATCT
101 CAGATGCCAT TGTTAATAAT TTCTTCCTTT CGGGTGGCGA GGTTGGTCAC
151 CAGAGACCAT CGGTCATTCC GCGAGCAACC CTGACCGCAA TGTTGACGTG
201 CTTGCGGCGT TTCCACCTTC ATGCCAAATA AGCGCCTTCC ATCAGCTTGC
250 TGAGGAGCTG GGCCACCGGC GGCGCCGGTG CTGTGCTACC TCCCTGCCCT
301 GAGCTTGAGC AGGGCCTTCT CTATCTGCCA CAGGAGCTAG CCTCCTGTGA
350 CAGTGTTGTG ACATTTGAGC TAACTGACAT TGTGCACTGC CGCATGGCGG
401 CCCCTAGCCA AAGGAAAGCT GTTTTGTCCA CGCTGGTAGG CCGGTATGGC
451 AGACGCACAA GGCTTTATGA TGCGGGTCAC ACCGATGTCC GCGCCTCCCT
501 TGCGCGCTTT ATTCCCACTC TCGGGCGGGT TACTGCCACC ACCTGTGAAC
551 TCTTTGAGCT TGTAGAGGCG ATGGTGGAGA AGGGCCAAGA CGGTTCAGCC
601 GTCCTCGAGT TGGATTTGTG CAGCCGAGAT GTCTCCCGCA TAACCTTTTT
651 CCAGAAGGAT TGTAACAAGT TCACGACCGG CGAGACAATT GCGCATGGCA
701 AAGTCGGTCA GGGTATCTCC GCCTGGAGTA AGACCTTTTG TGCCCTGTTT
751 GGCCCCTGGT TCCGTTGCGA TTGAGAAGGC TATTCTATCC CTTTTACCAC
801 AAGCTGTGTT CTACGGGGAT GCTTATGACG ACTCAGTATT CTCTGCTGCC
851 GTGGCTGGCG CCAGCCATGC CATGGTGTTT GAAAATGATT TTTCTGAGTT
901 TGACTCGACT CAGAATAACT TTTCCCTAGG TCTTGAGTGC GCCATTATGG
951 AAGAGTGTGG TATGCCCCAG TGGCTTGTCA GGTTGTACCA TGCCGTCCGG
1001 TCGGCGTGGA TCCTGCAGGC CCCAAAAGAG TCTTTGAGAG GGTTCTGGAA
1051 GAAGCATTCT GGTGAGCCGG GCACGTTGCT CTGGAATACG GTGTGGAACA
1101 TGGCAATCAT TGCCCATTGC TATGAGTTCC GGGACCTCCA GGTTGCCGCC
1151 TTCAAGGGCG ACGACTCGGT CGTCCTCTGT AGTGAATACC GCCAGAGCCC
1201 AGGCGCCGGT TCGCTTATAG CAGGCTGTGG TTTGAAGCCG AAGGCTGACT
1251 TCCGGCCGAT TGGGCTGTAT GCCGGGGTTG TCGTCGCCCC GGGGCTCGGG

1301 GCCCTACCCG ATGTCGTTCG ATTCGCCGGA CGGCTTTCGG AGAAGAACTG
1351 GGGGCCTGAT CCGGAGCGGG CAGAGCAGCT CCGCCTCGCC GTGCAGGATT
1401 TCCTCCGTAG GTTAACGAAT GTGGCCCAGA TTTGTGTTGA GGTGGTGTCT
1451 AGAGTTTACG GGGTTTCCCC GGGTCTGGTT CATAACCTGA TAGGCATGCT
1501 CCAGACTATT GGTGATGGTA AGGCGCATTT TACAGAGTCT GTTAAGCCTA
1551 TACTTGACCT TACACACTCA ATTATGCACC GGTCTGAATG AATAACATGT
1601 GGTTTCCTGC GCCCATGGGT TCGCCACCAT GCGCCCTAGG CCTCTTTTGC
1651 CGAGCGGCCG C,

or a sequence complementary thereto.

4. A protein which is (a) immunoreactive with antibodies present in individuals infected with enterically transmitted nonA/nonB and (b) derived from a viral hepatitis agent whose genome contains a region which is homologous to the 1.33 kb DNA EcoRI insert present in plasmid pTZ-KF1(ET1.1) carried in *E. coli* strain BB4, and having ATCC deposit no. 67717.

5. The protein of claim 4, which is encoded by a coding region within said 1.33 kb EcoRI insert.

6. A method of detecting infection by enterically transmitted nonA/nonB hepatitis viral agent in a test individual, comprising:

providing a peptide antigen which is (a) immunoreactive with antibodies present in individuals infected with enterically transmitted nonA/nonB hepatitis and (b) derived from a viral hepatitis agent whose genome contains a region which is homlogous to the 1.33 kb DNA EcoRI insert present in plasmid pTZ-KF1(ET1.1) carried in *E. coli* strain BB4, and having ATCC deposit no. 67717,

reacting serum from the test individual with such antigen, and

examining the antigen for the presence of

bound antibody.

7. The method of claim 6, wherein the serum antibody is an IgM or IgG antibody, or a mixture of both, the antigen provided is attached to a support, said reacting includes contacting such serum with the support and said examining includes reacting the support and bound serum antibody with a reporter-labeled anti-human antibody.

8. A kit for ascertaining the presence of serum antibodies which are diagnostic of enterically transmitted nonA/nonB hepatitis infection, comprising

a support with surface-bound recombinant peptide antigen which is (a) immunoreactive with antibodies present in individuals infected with enterically transmitted nonA/nonB viral hepatitis agent and (b) derived from a viral hepatitis agent whose genome contains a region which is homologous to the 1.33 kb DNA <u>Eco</u>RI insert present in plasmid pTZ-KF1(ET1.1) carried in <u>E. coli</u> strain BB4, and having ATCC deposit no. 67717,

a reporter-labeled anti-human antibody.

9. A DNA fragment derived from an enterically transmitted nonA/nonB viral hepatitis agent whose genome contains a region which is homologous to the 1.33 kb DNA <u>Eco</u>RI insert present in plasmid pTZ-KF1(ET1.1) carried in <u>E. coli</u> strain BB4 and having ATCC deposit no. 67717.

10. The fragment of claim 9, which is derived from said 1.33 kb <u>Eco</u>RI insert.

11. A DNA fragment derived from an enterically transmitted nonA/nonB viral hepatitis agent whose

genome contains a region which is homologous to a duplex DNA fragment within a first sequence:

```
1         11        21        31        41        51
*         *         *         *         *         *
CGAATTCCGCCAACTGATGGAAGGCACTAATCTGGCAAGACCTGTCCCTGTTGCAGCTGT

61        71        81        91        101       111
*         *         *         *         *         *
TCTACCACCCTGCCCCGAGCTCGAACAGGGCCTTCTCTACCTGCCCCAGGAGCTCACCAC

121       131       141       151       161       171
*         *         *         *         *         *
CTGTGATAGTGTCGTAACATTTGAATTAACAGACATTGTGCACTGCCGCATGGCCGCCCC

181       191       201       211       221       231
*         *         *         *         *         *
GAGCCAGCGCAAGGCCGTGCTGTCCACACTCGTGGGCCGCTACGGCGTCGCACAAAGCTC

241       251       261       271       281       291
*         *         *         *         *         *
TACAATGCTTCCCACTCTGATGTTCGCGACTCTCTCGCCCGTTTTATCCCGGCCATTGGC

301       311       321       331       341       351
*         *         *         *         *         *
CCCGTACAGGTTACAACTTGTGAATTGTACGAGCTAGTGGAGGCCATGGTCGAGAAGGGC

361       371       381       391       401       411
*         *         *         *         *         *
CAGGATGGCTCCGCCGTCCTTGAGCTTGATCTTTGCAACCGTGACGTGTCCAGGATCACC

421       431       441       451       461       471
*         *         *         *         *         *
TTCTTCCAGAAAGATTGTAACAAGTTCACCACAGGTGAGACCATTGCCCATGGTAAAGTG

481       491       501       511       521       531
*         *         *         *         *         *
```

GGCCAGGGCATCTCGGCCTGGAGCAAGACCTTCTGCGCCCTCTTTGGCCCTTGGTTCCGC

541 551 561 571 581 591
* * * * * *
GCTATTGAGAAGGCTATTCTGGCCCTGCTCCCTCAGGGTGTGTTTACGGTGATGCCTTTT

601 611 621 631 641 651
* * * * * *
GATGACACCGTCTTCTCGGCGGCTGTGGCCGCAGCAAAGGCATCCATGGTGTTTGAGAAT

661 671 681 691 701 711
* * * * * *
GACTTTTCTGAGTTTGACTCCACCCAGAATAACTTTTCTCTGGGTCTAGAGTGTGCTATT

721 731 741 751 761 771
* * * * * *
ATGGAGGAGTGTGGGATGCCGCAGTGGCTCATCCGCCTGTATCACCTTATAAGGTCTGCG

781 791 801 811 821 831
* * * * * *
TGGATCTTGCAGGCCCCGAAGGAGTCTCTGCGAGGGTTTTGGAAGAAACACTCCGGTGAG

841 851 861 871 881 891
* * * * * *
CCCGGCACTCTTCTATGGAATACTGTCTGGAATATGGCCGTTATTACCCACTGTTATGAC

901 911 921 931 941 951
* * * * * *
TTCCGCGATTTTCAGGTGGCTGCCTTTAAAGGTGATGATTCGATAGTGCTTTGCAGTGAG

961 971 981 991 1001 1011
* * * * * *
TATCGTCAGAGTCCAGGAGCTGCTGTCCTGATCGCCGGCTGTGGCTTGAAGTTGAAGGTA

1021 1031 1041 1051 1061 1071
* * * * * *
GATTTCCGCCCGATCGGTTTGTATGCAGGTGTTGTGGTGGCCCCCGGCCTTGGCGCGCTC

```
1081      1091      1101      1111      1121      1131
*         *         *         *         *         *
CCTGATGTTGTGCGCTTCGCCGGCCGGCTTACCGAGAAGAATTGGGGCCCTGGCCCTGAG

1141      1151      1161      1171      1181      1191
*         *         *         *         *         *
CGGGCGGAGCAGCTCCGCCTCGCTGTTAGTGATTTCCTCCGCAAGCTCACGAATGTAGCT

1201      1211      1221      1231      1241      1251
*         *         *         *         *         *
CAGATGTGTGTGGATGTTGTTTCCCGTGTTTATGGGGTTTCCCCTGGACTCGTTCATAAC

1261      1271      1281      1291      1301      1311
*         *         *         *         *         *
CTGATTGGCATGCTACAGGCTGTTGCTGATGGCAAGGCACATTTCACTGAGTCAGTAAAA

1321      1331      1341
*         *         *
CCAGTGCTCGACCGGAATTCGAGC,
```

a second sequence

```
1         11        21        31        41        51
*         *         *         *         *         *
GCTCGAATTCCGGTCGAGCACTGGTTTTACTGACTCAGTGAAATGTGCCTTGCCATCAGC

61        71        81        91        101       111
*         *         *         *         *         *
AACAGCCTGTAGCATGCCAATCAGGTTATGAACGAGTCCAGGGGAAACCCCATAAACACG

121       131       141       151       161       171
*         *         *         *         *         *
GGAAACAACATCCACACACATCTGAGCTACATTCGTGAGCTTGCGGAGGAAATCACTAAC

181       191       201       211       221       231
*         *         *         *         *         *
```

```
AGCGAGGCGGAGCTGCTCCGCCCGCTCAGGGCCAGGGCCCCAATTCTTCTCGGTAAGCCG

241       251       261       271       281       291
*         *         *         *         *         *
GCCGGCGAAGCGCACAACATCAGGGAGCGCGCCAAGGCCGGGGGCCACCACAACACCTGC

301       311       321       331       341       351
*         *         *         *         *         *
ATACAAACCGATCGGGCGGAAATCTACCTTCAACTTCAAGCCACAGCCGGCGATCAGGAC

361       371       381       391       401       411
*         *         *         *         *         *
AGCAGCTCCTGGACTCTGACGATACTCACTGCAAAGCACTATCGAATCATCACCTTTAAA

421       431       441       451       461       471
*         *         *         *         *         *
GGCAGCCACCTGAAAATCGCGGAAGTCATAACAGTGGGTAATAACGGCCATATTCCAGAC

481       491       501       511       521       531
*         *         *         *         *         *
AGTATTCCATAGAAGAGTGCCGGGCTCACCGGAGTGTTTCTTCCAAAACCCTCGCAGAGA
```

```
541       551       561       571       581       591
*         *         *         *         *         *
CTCCTTCGGGGCCTGCAAGATCCACGCAGACCTTATAAGGTGATACAGGCGGATGAGCCA

601       611       621       631       641       651
*         *         *         *         *         *
CTGCGGCATCCCACACTCCTCCATAATAGCACACTCTAGACCCAGAGAAAAGTTATTCTG

661       671       681       691       701       711
*         *         *         *         *         *
GGTGGAGTCAAACTCAGAAAAGTCATTCTCAAACACCATGGATGCCTTTGCTGCGGCCAC

721       731       741       751       761       771
*         *         *         *         *         *
AGCCGCCGAGAAGACGGTGTCATCAAAGGCATCACCGTAAAACACACCCTGAGGGAGCAG

781       791       801       811       821       831
*         *         *         *         *         *
GGCCAGAATAGCCTTCTCAATAGCGCGGAACCAAGGGCCAAAGAGGGCGCAGAAGGTCTT

841       851       861       871       881       891
*         *         *         *         *         *
GCTCCAGGCCGAGATGCCCTGGCCCACTTTACCATGGGCAATGGTCTCACCTGTGGTGAA

901       911       921       931       941       951
*         *         *         *         *         *
CTTGTTACAATCTTTCTGGAAGAAGGTGATCCTGGACACGTCACGGTTGCAAAGATCAAG

961       971       981       991       1001      1011
*         *         *         *         *         *
CTCAAGGACGGCGGAGCCATCCTGGCCCTTCTCGACCATGGCCTCCACTAGCTCGTACAA

1021      1031      1041      1051      1061      1071
*         *         *         *         *         *
TTCACAAGTTGTAACCTGTACGGGGCCAATGGCCGGGATAAAACGGGCGAGAGAGTCGCG
```

```
1081      1091      1101      1111      1121      1131
*         *         *         *         *         *
AACATCAGAGTGGGAAGCATTGTAGAGCTTTGTGCGACGCCGTAGCGGCCCACGAGTGTG

1141      1151      1161      1171      1181      1191
*         *         *         *         *         *
GACAGCACGGCCTTGCGCTGGCTCGGGGCGGCCATGCGGCAGTGCACAATGTCTGTTAAT

1201      1211      1221      1231      1241      1251
*         *         *         *         *         *
TCAAATGTTACGACACTATCACAGGTGGTGAGCTCCTGGGGCAGGTAGAGAAGGCCCTGT

1261      1271      1281      1291      1301      1311
*         *         *         *         *         *
TCGAGCTCGGGGCAGGGTGGTAGAACAGCTGCAACAGGGACAGGTCTTGCCAGATTAGTG

1321      1331      1341
*         *         *
CCTTCCATCAGTTGGCGGAATTCG,
```

a third sequence

```
  1 CGGTTGTTCA GTACCAGTTT ACTGCAGGTG TGCCTGGATC CGGCAAGTCC
 51 CGCTCTATCA CCCAAGCCGA TGTGGACGTT GTCGTGGTCC CGACGCGTGA
101 GTTGCGTAAT GCCTGGCGCC GTCGCGGCTT TGCTGCTTTT ACCCCGCATA
151 CTGCCGCCAG AGTCACCCAG GGGCGCCGGG TTGTCATTGA TGAGGCTCCA
201 TCCCTCCCCC CTCACCTGCT GCTGCTCCAC ATGCAGCGGG CCGCCACCGT
251 CCACCTTCTT GGCGCCCCGA ACCAGATCCC AGCCATCGAC TTTGAGCACG
301 CTGGGCTCGT CCCCGCCATC AGGCCCGACT TAGCCCCACC TCCTGGTGGC
351 ATGTTACCCA TCGCTGCCCT GCGGATGTAT GCGAGCTCAT CCGTGGTGCA
401 TACCCCATGA TCCAGACCAC TAGCCGGGTT CTCCGTTCGT TGTTCTGGGG
451 TGAGCCTGCC GTCGGGCAGA AACTAGTGTT CACCCAGGCG GCCAAGGCCG
501 CCAACCCCGG CTCAGTGACG GTCCACGAGG CGCAGGGCGC TACCTACACG
551 GAGACCACTA TTATTGCCAC AGCAGATGCC CGGGGCCTTA TTCAGTCGTC
601 TCGGGCTCAT GCCATTGTTG CTCTGACGCG CCACACTGAG AAGTGCGTCA
651 TCATTGACGC ACCAGGCCTG CTTCGCGAGG TGGGCATCTC CGATGCAATC
701 GTTAATAACT TTTTCCTCGC TGGTGGCGAA ATTGGTCACC AGCGCCCATC
```

751 AGTTATTCCC CGTGGCAACC CTGACGCCAA TGTTGACACC CTGGCTGCCT
801 TCCCGCCGTC TTGCCAGATT AGTGCCTTCC ATCAGTTGGC TGAGGAGCTT
851 GGCCACAGAC CTGTCCCTGT TGCAGCTGTT CTACCACCCT GCCCCGAGCT
901 CGAACAGGGC CTTCTCTACC TGCCCCAGGA GCTCACCACC TGTGATAGTG
951 TCGTAACATT TGAATTAACA GACATTGTGC ACTGCCGCAT GGCCGCCCCG
1001 AGCCAGCGCA AGGCCGTGCT GTCCACACTC GTGGGCCGCT ACGGCGTCGC
1051 ACAAAGCTCT ACAATGCTTC CCACTCTGAT GTTCGCGACT CTCTCGCCCG
1101 TTTTATCCCG GCCATTGGCC CCGTACAGGT TACAACTTGT GAATTGTACG
1151 AGCTAGTGGA GGCCATGGTC GAGAAGGGCC AGGATGGCTC CGCCGTCCTT
1201 GAGCTTGATC TTTGCAACCG TGACGTGTCC AGGATCACCT TCTTCCAGAA
1251 AGATTGTAAC AAGTTCACCA CAGGTGAGAC CATTGCCCAT GGTAAAGTGG
1301 GCCAGGGCAT CTCGGCCTGG AGCAAGACCT TCTGCGCCCT CTTTGGCCCT
1351 TGGTTCCGCG CTATTGAGAA GGCTATTCTG GCCCTGCTCC CTCAGGGTGT
1401 GTTTTACGGT GATGCCTTTG ATGACACCGT CTTCTCGGCG GCTGTGGCCG
1451 CAGCAAAGGC ATCCATGGTG TTTGAGAATG ACTTTTCTGA GTTTGACTCC
1501 ACCCAGAATA ACTTTTCTCT GGGTCTAGAG TGTGCTATTA TGGAGGAGTG
1551 TGGGATGCCG CAGTGGCTCA TCCGCCTGTA TCACCTTATA AGGTCTGCGT
1601 GGATCTTGCA GGCCCCGAAG GAGTCTCTGC GAGGGTTTTG GAAGAAACAC
1651 TCCGGTGAGC CCGGCACTCT TCTATGGAAT ACTGTCTGGA ATATGGCCGT
1701 TATTACCCAC TGTTATGACT TCCGCGATTT TCAGGTGGCT GCCTTTAAAG
1751 GTGATGATTC GATAGTGCTT TGCAGTGAGT ATCGTCAGAG TCCAGGAGCT
1801 GCTGTCCTGA TCGCCGGCTG TGGCTTGAAG TTGAAGGTAG ATTTCCGCCC
1851 GATCGGTTTG TATGCAGGTG TTGTGGTGGC CCCCGGCCTT GGCGCGCTCC
1901 CTGATGTTGT GCGCTTCGCC GGCCGGCTTA CCGAGAAGAA TTGGGGCCCT
1951 GGCCCTGAGC GGGCGGAGCA GCTCCGCCTC GCTGTTAGTG ATTTCCTCCG
2001 CAAGCTCACG AATGTAGCTC AGATGTGTGT GGATGTTGTT TCCCGTGTTT
2051 ATGGGGTTTC CCCTGGACTC GTTCATAACC TGATTGGCAT GCTACAGGCT
2101 GTTGCTGATG GCAAGGCACA TTTCACTGAG TCAGTAAAAC CAGTGCTCGA
2151 CTTGACAAAT TCAATCTTGT GTCGGGTGGA ATGAATAACA TGTCTTTTGC
2201 TGCGCCCATG GGTTCGCGAC CATGCGCCCT CGGCCTATTT TGTTGCTGCT
2251 CCTCATGTTT TTGCCTATGC TGCCCGCGCC ACCGCCCGGT CAGCCGTCTG
2301 GCCGCCGTCG TGGGCGGCGC AGCGGCGGTT CCGGCGGTGG TTTCTGGGGT
2351 GACCGGGTTG ATTCTCAGCC CTTCGCAATC CCCTATATTC ATCCAACCAA
2401 CCCCTTCGCC CCCGATGTCA CCGCTGCGGC CGGGGCTGGA CCTCGTGTTC
2451 GCCAACCCGC CCGACCACTC GCGTCCGCTT GGCGTGACCA GGCCCAGCGC
2501 CCCGCCGTTG CCTCACGTCG TAGACCTACC ACAGCTGGGG CCGCGCCGCT
2551 AACCGCGGTC GCTCCGGCCC CG,

or a fourth sequence

```
   1 GCGGCCGCTC GGTGGGGTTC CACCTCTCTT CCCATTCCGA ACAGAGAAGT
  51 TAAGCAAGGA ATTAATTCGC GGCCGCTCGT GTTGCGTGAG GTGGGTATCT
 101 CAGATGCCAT TGTTAATAAT TTCTTCCTTT CGGGTGGCGA GGTTGGTCAC
 151 CAGAGACCAT CGGTCATTCC GCGAGCAACC CTGACCGCAA TGTTGACGTG
 201 CTTGCGGCGT TTCCACCTTC ATGCCAAATA AGCGCCTTCC ATCAGCTTGC
 250 TGAGGAGCTG GGCCACCGGC GGCGCCGGTG CTGTGCTACC TCCCTGCCCT
 301 GAGCTTGAGC AGGGCCTTCT CTATCTGCCA CAGGAGCTAG CCTCCTGTGA
 350 CAGTGTTGTG ACATTTGAGC TAACTGACAT TGTGCACTGC CGCATGGCGG
 401 CCCCTAGCCA AAGGAAAGCT GTTTTGTCCA CGCTGGTAGG CCGGTATGGC
 451 AGACGCACAA GGCTTTATGA TGCGGGTCAC ACCGATGTCC GCGCCTCCCT
 501 TGCGCGCTTT ATTCCCACTC TCGGGCGGGT TACTGCCACC ACCTGTGAAC
 551 TCTTTGAGCT TGTAGAGGCG ATGGTGGAGA AGGGCCAAGA CGGTTCAGCC
 601 GTCCTCGAGT TGGATTTGTG CAGCCGAGAT GTCTCCCGCA TAACCTTTTT
 651 CCAGAAGGAT TGTAACAAGT TCACGACCGG CGAGACAATT GCGCATGGCA
 701 AAGTCGGTCA GGGTATCTCC GCCTGGAGTA AGACCTTTTG TGCCCTGTTT
 751 GGCCCCTGGT TCCGTTGCGA TTGAGAAGGC TATTCTATCC CTTTTACCAC
 801 AAGCTGTGTT CTACGGGGAT GCTTATGACG ACTCAGTATT CTCTGCTGCC
 851 GTGGCTGGCG CCAGCCATGC CATGGTGTTT GAAAATGATT TTTCTGAGTT
 901 TGACTCGACT CAGAATAACT TTTCCCTAGG TCTTGAGTGC GCCATTATGG
 951 AAGAGTGTGG TATGCCCCAG TGGCTTGTCA GGTTGTACCA TGCCGTCCGG
1001 TCGGCGTGGA TCCTGCAGGC CCCAAAAGAG TCTTTGAGAG GGTTCTGGAA
1051 GAAGCATTCT GGTGAGCCGG GCACGTTGCT CTGGAATACG GTGTGGAACA
1101 TGGCAATCAT TGCCCATTGC TATGAGTTCC GGGACCTCCA GGTTGCCGCC
1151 TTCAAGGGCG ACGACTCGGT CGTCCTCTGT AGTGAATACC GCCAGAGCCC
1201 AGGCGCCGGT TCGCTTATAG CAGGCTGTGG TTTGAAGCCG AAGGCTGACT
1251 TCCGGCCGAT TGGGCTGTAT GCCGGGGTTG TCGTCGCCCC GGGGCTCGGG
1301 GCCCTACCCG ATGTCGTTCG ATTCGCCGGA CGGCTTTCGG AGAAGAACTG
1351 GGGGCCTGAT CCGGAGCGGG CAGAGCAGCT CCGCCTCGCC GTGCAGGATT
1401 TCCTCCGTAG GTTAACGAAT GTGGCCCAGA TTTGTGTTGA GGTGGTGTCT
1451 AGAGTTTACG GGGTTTCCCC GGGTCTGGTT CATAACCTGA TAGGCATGCT
1501 CCAGACTATT GGTGATGGTA AGGCGCATTT TACAGAGTCT GTTAAGCCTA
1551 TACTTGACCT TACACACTCA ATTATGCACC GGTCTGAATG AATAACATGT
1601 GGTTTCCTGC GCCCATGGGT TCGCCACCAT GCGCCCTAGG CCTCTTTTGC
1651 CGAGCGGCCG C,
```

or a sequence complementary thereto.

12. The DNA fragment of claim 11, wherein said fragment contains a coding sequence homologous to nucleotides 2 through 101 of said first sequence, nucleotides 594 through 643 of said second sequence, or a sequence complementary to said coding sequences.

13. A kit comprising, in a container or separate containers, a pair of single-strand primers derived from non-homologous regions of opposite strands of a DNA duplex fragment derived from an enterically transmitted viral hepatitis agent whose genome contains a region which is homologous to the 1.33 kb DNA EcoRI insert present in plasmid pTZ-KF1(ET1.1) carried in *E. coli* strain BB4 and having ATCC deposit no. 67717.

14. The kit of claim 13, which are derived from opposite strands of the EcoRI duplex insert in said plasmid.

15. A method for detecting the presence of an enterically transmitted nonA/nonB hepatitis viral agent in a biological sample, comprising

preparing a mixture of duplex DNA fragments derived from the sample,

denaturing the duplex fragments,

adding to the denatured DNA fragments, a pair of single-strand primers derived from non-homologous regions of opposite strands of a DNA duplex fragment derived from an enterically transmitted viral hepatitis agent whose genome contains a region which is homologous to the 1.33 kb DNA EcoRI insert present in plasmid pTZ-KF1(ET1.1) carried in *E. coli* strain BB4, and having ATCC deposit no. 67717,

hybridizing said primers to homologous-

sequence region of opposite strands of such duplex DNA fragments derived from enterically transmitted nonA/nonB hepatitis agent,

reacting the primed fragment strands with DNA polymerase in the presence of DNA nucleotides, to form new DNA duplexes containing the primer sequences, and

repeating said denaturing, adding, hybridizing and reacting steps, until a desired degree of amplification of sequences is achieved.

16. The method of claim 15, wherein the primers are derived from opposite strands the EcoRI duplex insert in said plasmid.

17. The method of claim 15, for detecting the presence of viral agent in a sample of cultured cells infected with the agent.

18. A vaccine for immunizing an individual against enterically transmitted nonA/nonB hepatitis viral agent comprising, in a pharmacologically acceptable adjuvant, a recombinant protein derived from an enterically transmitted nonA/nonB viral hepatitis agent whose genome contains a region which is homologous to the 1.33 kb DNA EcoRI insert present in plasmid pTZ-KF1(ET1.1) carried in E. coli strain BB4, and having ATCC deposit no. 67717.

19. The vaccine of claim 18, wherein the protein is derived from the EcoRI insert in said plasmid.

20. In a method of isolating an enterically transmitted non-A/nonB viral agent or a nucleic acid fragment produced by the agent, an improvement which comprises:

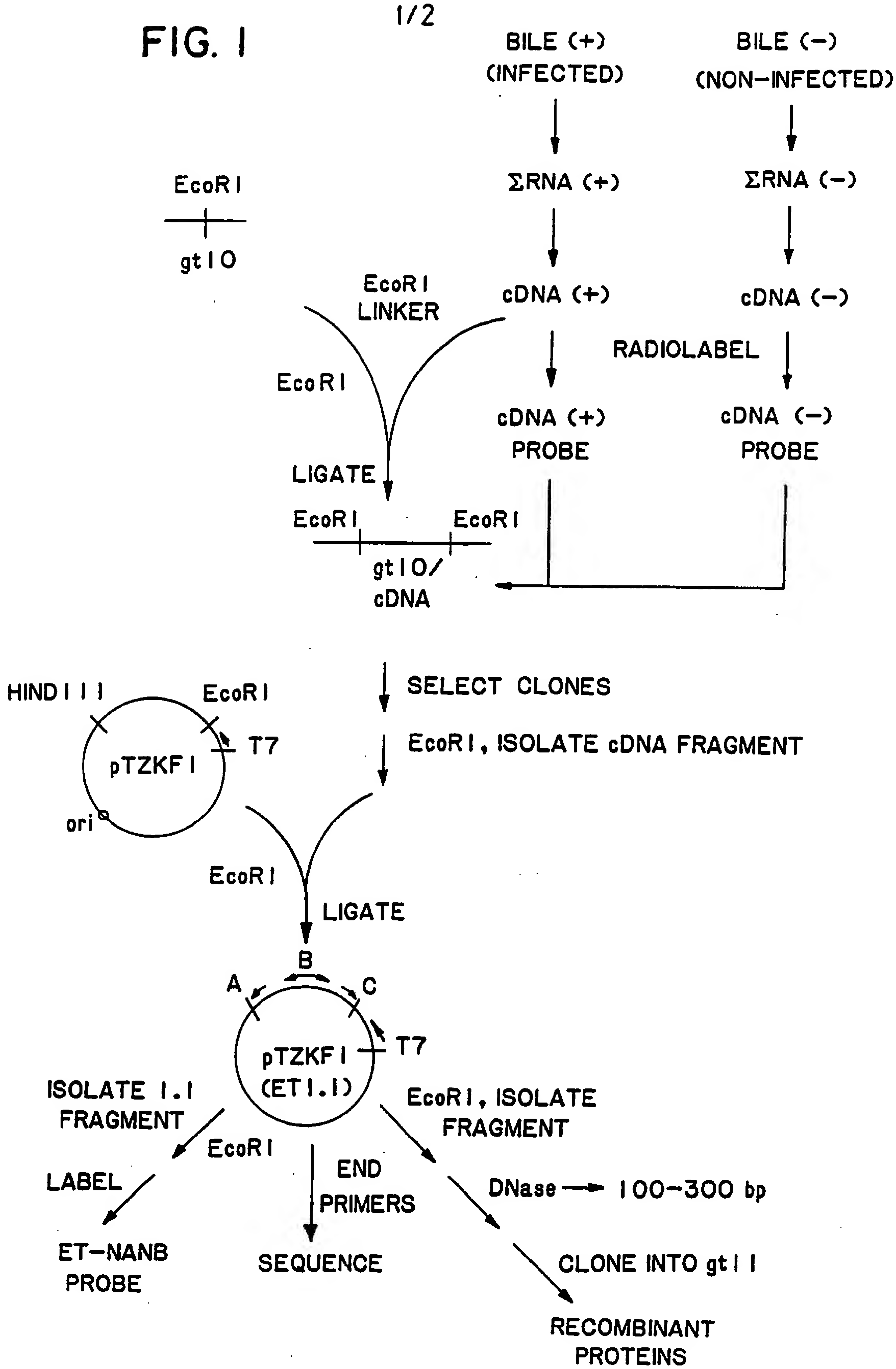
utilizing, as a source of said agent, bile obtained from a human or cynomolgus monkey having an active infection of enterically transmitted non-A/non-B hepatitis.

21. The method of claim 20, wherein the bile is obtained from an infected cynomologus monkey.

22. Human polyclonal anti-serum obtained from a human immunized with a protein derived form an enterically transmitted non-A/non-B viral hepatitis agent whose genome contains a region which is homologous to the 1.33 kb DNA <u>Eco</u>RI insert present in plasmid pTZ-KF1(ET1.1) carried in <u>E. coli</u> strain BB4 and having ATCC deposit no. 67717.

1/2

FIG. 1


BILE (+)
(INFECTED)
BILE (−)
(NON−INFECTED)
ΣRNA (+)
ΣRNA (−)
EcoR I
gt l O
EcoR I
LINKER
cDNA (+)
cDNA (−)
RADIOLABEL
EcoR I
cDNA (+)
PROBE
cDNA (−)
PROBE
LIGATE
EcoR I
EcoR I
gt l O/
cDNA
HIND I I I
EcoR I
T7
pTZKF I
ori
SELECT CLONES
EcoR I, ISOLATE cDNA FRAGMENT
EcoR I
LIGATE
B
A
C
T7
pTZKF I
(ET I.I)
ISOLATE I.I
FRAGMENT
EcoR I
EcoR I, ISOLATE
FRAGMENT
LABEL
END
PRIMERS
DNase → I 00−300 bp
ET−NANB
PROBE
SEQUENCE
CLONE INTO gt I I
RECOMBINANT
PROTEINS

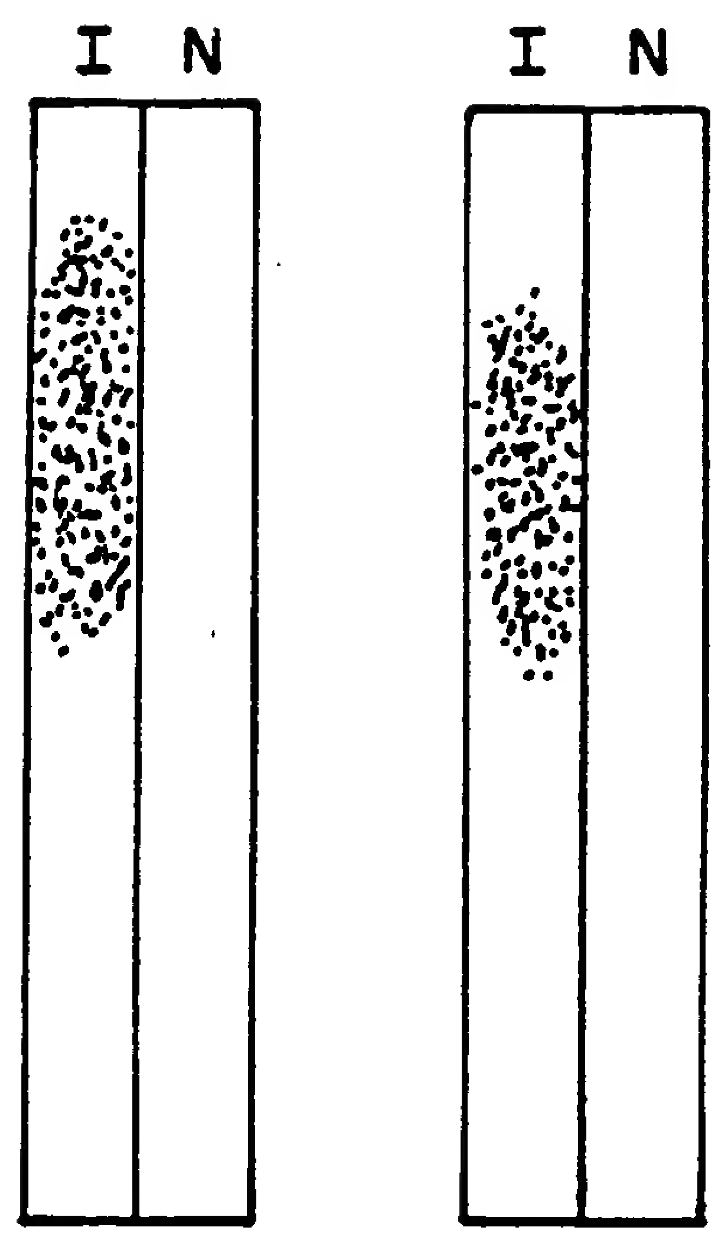


FIG. 2A FIG. 2B

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US89/02648

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [8]

According to International Patent Classification (IPC) or to both National Classification and IPC

INT. CL. 4 A61K 39/29,42; CO7H 17/00; CO7K 7/10, 15/04;See
U.S. CL. 424/86,89; 435/5,6,239,810; 436/513, 518, 536,Attach.

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
INT. CL. U.S. CL.	A61K 39/29,42 424/86,89; 435/5,6,239,810; 436/513,518, See Attach

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

APS AND BIOSIS SEARCHES: (ET OR ENTERIC?) (5A) (NANB? OR NON (W) B)

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
Y	WO, A, 85/01517 (MASSACHUSETTS INSTITUTE OF TECHNOLOGY) 30 September 1983. See page 5, line 6-page 6, line 27; page 14, lines 10-27; page 15, line 18-page 16, line 24.	1-19, 22
Y	US,A 4,591,552 (NEURATH) 29 September 1982 See col. 15, lines 22-29; col. 19, lines 58-68.	6-8
X Y	The Lancet, vol i, no. 8585, issued 12 March 1988, V.A. ARANKALLE ET AL, "Aetiological associated of a virus like particle with enterically transmitted non-A, non-B hepatitis", pages 550-554. See Summary and page 550, col. 2, 4th paragraph.	20,22 1-19

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
08 August 1989	17 OCT 1989
International Searching Authority	Signature of Authorized Officer
ISA/US	David A. Saunders David A. Saunders

Form PCT/ISA/210 (second sheet) (Rev.11-87)

International Application No. PCT/US89/02648

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
X Y	Journal of General Virology, vol. 69, no. 3 issued March 1988, D. BRADLEY ET AL., "Aetiological agent of enterically transmitted non-A non-B hepatitis", pages 731-738, See Summary and page 732, 2nd paragraph.	20-22 1-19
X Y	Proceedings of the National Academy of Sciences USA, vol. 84 no. 17, issued September 1987, D.W. BRADLEY ET AL, "Enterically transmitted non-A, non-B hepatitis. Serial passage of disease in cynomolgus macaques and tamarins and recovery of disease-associated 27-to 34-NM viruslike particles", pages 6277-6281. See abstract, paragraph spanning pages 6277-6278, and col. 1 of Table 2.	20-22 1-19
P,X P,Y	The Journal of Infections Disease, vol.159, no. 6, issued June 1989, K. KRAWCZYNSKI ET AL, "Enterically transmitted non-A, non-B, hepatitis: Indentitication of virus-assocaited antigen in experimentally infected cynomolgus macques", pages 1042-1049. Cols. 1 and 2; page 1048 Col. 2.	20-22 1-19
Y	US,A 4,683,202 (MULLIS) 28 July 1987. See Summary of the Invention; col. 6, lines 7-44; col. 14, line 66-col. 15,line 15.	15-17

International Application No. PCT/US89/02648

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

V.☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE [1]

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1.☐ Claim numbers , because they relate to subject matter [12] not required to be searched by this Authority, namely:

2.☐ Claim numbers , because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out [13], specifically:

3.☐ Claim numbers__________, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI.☒ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING [2]

This International Searching Authority found multiple inventions in this international application as follows:

See Attachment No. 2

1.☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application. Telephone practice

2.☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3.☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4.☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

ATTACHMENT

PCT/US89/02648

Attachment to PCT/ISA/210
International Search Report

I. Classification of Subject Matter.

INT. CL. Cont. C12N 7/02; C12Q 1/68,70; GOIN 33/531, 536, 543, 576.

U.S. CL. Cont. 543, 820; 530/324, 387, 806, 808, 826

II. Fields Searched

U.S. CL. Cont. 436/536, 543, 820; 530/324, 387, 806, 808, 826